مجله دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران (پیاورد سلامت)
دوره ۱۱ شماره ۳ مرداد و شهریور ۱۳۹۶، ۲۹۶-۲۸۷

مقاله پژوهشی

پیش بینی روش درمان بیماری قلبی با استفاده از الگوریتم های داده کاوی

سجاد مظاهری[۱]، مریم عاشوری[۲]، زینب بچاری[۳]

دریافت مقاله : آذر ۱۳۹۵
پذیرش مقاله : فروردین ۱۳۹۶

*نویسنده مسئول :
سجاد مظاهری؛
اداره کل بیمه سلامت استان سیستان و بلوچستان

Email :
sajad.mazaheri@gmail.com

چکیده

زمینه و هدف: امروزه بیماری قلبی بسیار معمول و یکی از دلایل اصلی مرگ و میر به شمار می رود. به همین علت تشخیص درست و به موقع این بیماری بسیار مهم است. روش های تشخیصی و درمانی این بیماری عوارض جانبی زیاد و پرهزینه ای دارد. بنابراین محققان به دنبال روش های ارزان و با دقت بالا برای تشخیص این بیماری هستند. پژوهش حاضر با هدف شناسایی الگویی برای تشخیص روش درمان بیماری قلبی صورت گرفته است.

روش بررسی: پژوهش حاضر به روش توصیفی- مقطعی صورت گرفته و نمونه گیری به روش سرشماری بوده است. جامعه ی پژوهش متشکل از داده های بیمارستان های خاتم الانبیاء(ع) و علی ابن ابی طالب(ع) زاهدان است که با مراجعه ی مستقیم پژوهش گر به سازمان حاصل و به صورت فایل اکسل تهیه گردید. جهت تحلیل داده ها از نرم افزار **Clementine12.0** استفاده شده است. در پژوهش حاضر الگوریتم های **C5.0**، **C&R Tree**، **CHAID** و **QUEST** و شبکه عصبی مصنوعی روی مجموعه داده اجرا گردید.

یافته ها: مقدار صحت ۷۶/۰۴ توسط الگوریتم **C&R Tree** نشان دهنده ی عملکرد بهتر الگوریتم های درخت تصمیم نسبت به شبکه ی عصبی است.

نتیجه گیری: هدف این مطالعه ارایه مدلی برای پیش بینی روش درمانی مناسب بیماری قلبی به منظور کاهش هزینه های درمان و کیفیت ارایه خدمات بهتر به پزشکان می باشد. با توجه به قابل ملاحظه بودن خطرات اجرای روش های تشخیص تهاجمی مانند آنژیوگرافی و نیز حصول تجارب موفقیت آمیز داده کاوی در پزشکی، این مطالعه مدلی مبتنی بر تکنیک های داده کاوی ارایه نموده است. نقطه ی قابل بهبود مدل فوق ارایه سیستمی تصمیم یار جهت کمک به پزشکان برای افزایش صحت تشخیص روش درمان بیماری می باشد.

واژه های کلیدی: داده کاوی، بیماری قلبی، پیش بینی، روش درمان، درخت تصمیم، شبکه عصبی

[۱] کارشناس ارشد مهندسی کامپیوتر- نرم افزار، اداره کل بیمه سلامت استان سیستان و بلوچستان، سازمان بیمه سلامت ایران، زاهدان، ایران

[۲] مربی گروه فناوری اطلاعات، مجتمع آموزش عالی سراوان، سراوان، ایران

[۳] کارشناس پرستاری، اداره کل بیمه سلامت استان سیستان و بلوچستان، سازمان بیمه سلامت ایران، زاهدان، ایران

[Downloaded from journals.tums.ac.ir on 2024-11-17]

مقدمه

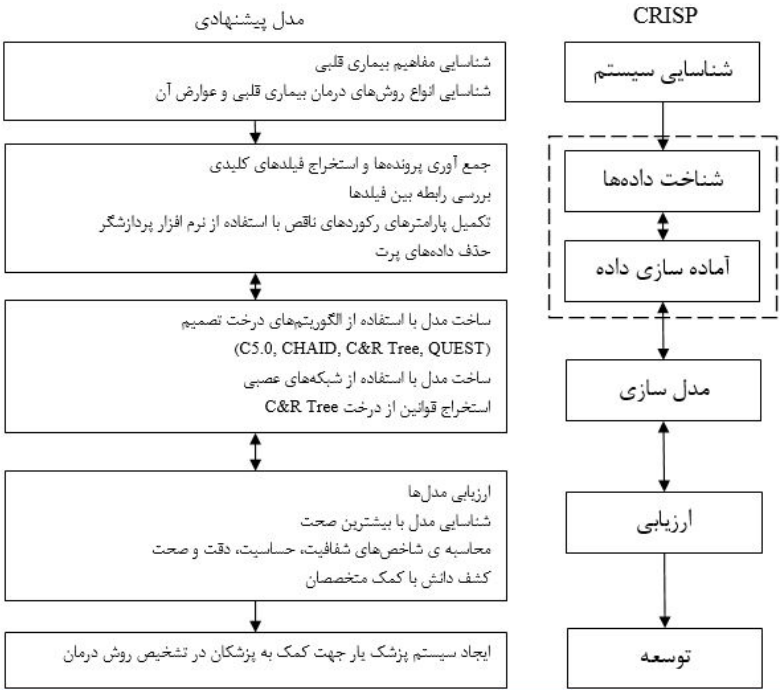
زندگی انسان ها وابسته به عملکرد مناسب قلب است و اگر عملکرد قلب به صورت مناسب صورت نگیرد روی سایر قسمت های بدن از قبیل ذهن، کلیه و غیره اثر خواهد گذاشت(۱). بیماری های قلبی و عروقی مهم ترین عامل مرگ و میر در دنیا می باشند. طبق اعلام سازمان جهانی بهداشت ۱۲ میلیون مرگ در کل دنیا بر اثر بیماری های قلبی صورت می گیرد. بیشترین علت مرگ و میرها در ایران ناشی از ابتلا به بیماری قلبی و عروقی است و مرگ ۳۸ درصد ایرانیان بر اثر بیماری قلبی رقم بسیار بالایی محسوب می‌شود(۲).

طبق نظر کارشناسان، دانش نوین داده کاوی از جمله دانش های در حال توسعه است که در سال های اخیر در تمامی عرصه ها جایگاه خود را تثبیت کرده است؛ به گونه ای که رشد آن در مقایسه با سایر دانش های برتر، فزاینده است. همکاری متخصصان در زمینه ی کامپیوتر و پزشکی راه حل جدیدی را در تحلیل داده های پزشکی و به دست آوردن الگوهای مفید و کاربردی ارایه می دهد که همان داده کاوی پزشکی است. حوزه ی پزشکی و سلامت از بخش های مهم در جوامع صنعتی است. استخراج دانایی از میان حجم انبوه داده های مرتبط با سوابق بیماری و پرونده های پزشکی افراد با استفاده از فرایند داده کاوی می تواند منجر به شناسایی قوانین حاکم بر ایجاد، رشد و تسری بیماری ها شده و اطلاعات ارزشمندی را به منظور شناسایی علل رخداد بیماری ها، تشخیص، پیش بینی و درمان بیماری ها با توجه به عوامل محیطی حاکم در اختیار متخصصان و دست اندرکاران حوزه ی سلامت قرار دهد. نتیجه ی این مسئله به معنای افزایش عمر و ایجاد آرامش برای افراد جامعه است. بنابراین پیچیدگی اطلاعات پزشکی و وجود ابزارهای داده کاوی باعث می شود که داده کاوی بر روی داده های پزشکی و سلامت مهم تلقی گردد(۳).

مهمترین خدمات قابل ارایه با استفاده از روش های داده کاوی عبارتند از:

- بررسی میزان تاثیر دارو بر بیماری و اثرات جانبی آن
- تشخیص و پیش بینی انواع بیماری ها مانند تشخیص یا پیش بینی انواع سرطان
- تعیین روش درمان بیماری ها
- پیش بینی میزان موفقیت اقدامات پزشکی مانند اعمال جراحی
- تجزیه و تحلیل داده های موجود در سیستم های اطلاعات سلامت(Health Information System: HIS)
- تحلیل عکس های پزشکی

یکی از روش های بسیار قوی برای پیاده سازی و اجرای پروژه های داده کاوی متدولوژی(CRISP یا Cross Industry Process for Data Mining) است(۴). در مقاله ی حاضر مدل پیشنهادی بر اساس CRISP که شامل پنج فاز است، ارایه شده است. هر یک از این فازها خود شامل زیر بخش هایی می شوند. حرکت رو به جلو و عقب بین فازهای مختلف نیاز است، زیرا ورودی هر فاز به خروجی فاز مرحله ی قبل وابسته است(۵). هر یک از این پنج فاز در شکل ۱ نشان داده شده اند.



شکل ۱: گام های روشCRISP و مدل پیشنهادی

[Downloaded from journals.tums.ac.ir on 2024-11-17]

مطالعات موجود در زمینه ی ترکیب دو حوزه ی داده کاوی و بیماری قلب موید این مطلب است که Palaniappan و Awang سیستم پیش بینی بیماری به نام قلب هوشمند را با استفاده از تکنیک های داده کاوی درخت تصمیم، شبکه بیز و شبکه عصبی توسعه داده اند. این سیستم به سوالات پیچیده ی پزشکی بر مبنای اگر و آنگاه پاسخ می دهد در حالی که سیستم های سنتی نمی توانند به این سوالات پاسخ دهند. سیستم مورد نظر مبتنی بر وب، کاربرپسند، مقیاس پذیر، قابل اعتماد و قابل ارتقا بوده و با استفاده از داده کاوی روی پارامترهای پزشکی از قبیل: سن، جنسیت، فشار خون، قند خون و غیره می توان احتمال بیمارشدن به بیماری های قلبی را پیش بینی کرد(۶). Venkatalakshmi و Shivsankar به طراحی و تشخیص سیستمی برای پیش بینی بیماری قلبی مبتنی بر تکنیک های داده کاوی پرداختند. مقدار ۸۴/۰۱ درصد برای صحت مدل ایجاد شده با درخت تصمیم حاصل گردید(۷). Masethe و Masethe به ارایه سیستمی برای پیش بینی حمله قلبی با صحت ۹۹٪ پرداختند. پارامترهای مورد بررسی شامل: جنسیت، سن، نوع درد قفسه سینه، افزایش ضربان قلب، کلسترول، سیگاری بودن، قند خون، فشارخون، دستگاه ثبت ضربان قلب، رژیم غذایی و الکل بود(۸).

Srinivas و همکاران پیش بینی بیماری قلبی را با تکنیک دسته بندی و الگوریتم های شبکه ی عصبی، شبکه ی بیزی، درخت تصمیم و ماشین بردار پشتیبان انجام دادند. شبکه بیزی وابستگی های شرطی بین متغیرها را شرح می دهد. بررسی نتایج پژوهش وی نشان می دهد که درخت تصمیم نسبت به سایر الگوریتم ها از دقت بالاتری برخوردار بوده است(۹). Soni و همکاران به بررسی عملکرد الگوریتم های مختلف داده کاوی با هدف پیش بینی حمله قلبی موثر پرداختند. نتایج نشان داد که درخت های تصمیم و شبکه ی بیزی صحت برابر و بهتر نسبت به سایر الگوریتم ها چون شبکه ی عصبی داشتند(۱۰). Tanejaaneja با استفاده از الگوریتم های داده کاوی به ارایه سیستمی برای پیش بینی حمله ی قلبی پرداخت. وی سه الگوریتم J48، شبکه بیزی و شبکه عصبی را به کار برد که الگوریتم J48 هرس شده با میزان صحت ۹۵/۵۶٪ و میزان دقت ۹۵/۵٪ بهترین مدل را ایجاد نمود(۱۱).

رحیمی شاطرانلو و علیزاده با ارایه مدل ترکیبی درخت تصمیم- شبکه بیزی به پیش بینی بیماری کرونری قلبی پرداختند. دقت و صحت مدل ترکیبی برابر ۹۵٪ و ۹۵٪ بود که بر مدل اولیه شبکه بیز با صحت ۸۹٪ و دقت ۹۰٪ برتری داشت(۱۲). صفدری و همکاران از الگوریتم های شبکه عصبی و درخت تصمیم برای پیشگویی ابتلا به آنفارکتوس قلبی استفاده نموده اند. صحت ۹۳/۴ درصد برای درخت تصمیم C5.0 نشان دهنده ی عملکرد بهتر این الگوریتم نسبت به سایر الگوریتم های درخت تصمیم و شبکه عصبی روی مجموعه داده ی مورد بررسی است(۱۳).

علیزاده ثانی به ارایه سیستمی برای تشخیص بیماری های قلبی روی ۳۰۳ نمونه(۸۷ نمونه سالم و ۲۱۶ نمونه بیمار) پرداخت(۱۴). محمودی و همکاران نیز به پیش بینی عروق کرونر با استفاده از شبکه های عصبی و گزینش متغیر مبتنی بر درخت رگرسیون و طبقه بندی پرداختند. مقادیر ۷۴/۱۹، ۳۳/۲۵ و ۹۲/۴۱ برای شاخص های دقت، ویژگی و حساسیت حاصل گردید(۱۵).

امروزه بسیاری از بیمارستان ها از سیستم مدیریت اطلاعات بیمارستان برای مدیریت داده های بیماران و یا بهداشت و درمان استفاده می کنند. این سیستم ها مقدار زیادی داده شامل اعداد، متون، نمودارها و عکس ها ایجاد می کنند. متاسفانه این داده ها به ندرت در تصمیم گیری های کلینیکی مورد استفاده قرار می گیرند. سوالی که ذهن اکثر متخصصان این حوزه را به خود مشغول کرده این است که چگونه می توان داده ها را تبدیل به اطلاعات مفید نمود که کادر پزشکی را قادر به تصمیمات کلینیکی هوشمندانه نماید. از سوی دیگر، روش های داده کاوی روی داده ها به دنبال تایید آنچه از قبل وجود دارد، نیستند بلکه به دنبال مشخص کردن الگوهای از قبل شناخته نشده هستند. همچنین هزینه و عوارض جانبی روش های تشخیصی بیماری قلب سبب شد که محققان به دنبال روش های ارزان و با دقت بالا برای تشخیص این بیماری باشند.

بنابراین پژوهش حاضر به دنبال پاسخگویی به این قبیل سوال هاست:

- چگونه می توان داده های موجود را به اطلاعات مفید برای اتخاذ تصیمات کلنیکی هوشمند تبدیل نمود؟
- چگونه می توان از داده های موجود، الگوهایی را که از قبل شناخته نشده اند، شناسایی نمود؟
- چگونه می توان روش هایی با دقت بالا و ارزان قیمت برای تشخیص و درمان بیماری ها و به ویژه بیماری قلبی ابداع نمود؟

داده کاوی به عنوان یک حوزه ی نوظهور با پردازش داده ها و اطلاعات موجود به کشف دانش و شناسایی الگوهای جدید می پردازد. استفاده از تکنیک های مختلف داده کاوی، مدل هایی با دقت بالا ارایه می دهند که می توانند به عنوان سیستم های تصمیم یار استفاده شوند.

وجود رویکردهایی جهت پیش بینی بیماری قلبی و نبود سیستمی جهت تشخیص روش درمان بیماری سبب گردید که در

[Downloaded from journals.tums.ac.ir on 2024-11-17]

مطالعه ی حاضر به شناسایی الگویی برای تشخیص روش درمان بیماری های قلبی پرداخته شود. در این پژوهش، هدف، شناسایی افراد بیمار قلبی نیست بلکه تشخیص روش درمان مناسب مورد توجه است. بنابراین روش دسته بندی مورد توجه قرار گرفت و الگوریتم های C5.0، C&R Tree، CHAID و QUEST و شبکه عصبی مصنوعی روی مجموعه داده اجرا گردید. مقدار صحت الگوریتم C&R Tree نشان دهنده ی عملکرد بهتر الگوریتم های درخت تصمیم نسبت به شبکه عصبی می باشد.

روش بررسی

الف) شناخت سیستم

در این مرحله به شناخت سیستم مورد نظر پرداخته می شود و سپس اهداف مورد نظر و عوامل موفقیت کلیدی سیستم تعیین و بازنگری می گردد. طبق نظر متخصصان قلب، با توجه به رشد روز افزون بیماری های قلبی، هزینه های سرسام آور درمان این بیماری و عوارض شدیدی که روی اعضای حیاتی بدن در دراز مدت می گذارد، بررسی داده های جمع آوری شده در رابطه با این بیماری برای تشخیص روش درمان بیماران جدید می تواند مفید باشد. بیماران جدید می توانند تا حد ممکن از توصیه های پزشکی تجویز شده متناسب با بیماران دسته ای که در آن قرار گرفته اند، بهره ببرند. همچنین از رژیم غذایی و نوع برنامه غذایی تجویز شده برای آن بیماران استفاده کنند تا شدت بیماری آن ها تا حدودی مهار گردد.

ب) شناخت داده ها و آماده سازی آن ها

در این فاز به جمع آوری داده های اولیه، توصیف داده ها، بازرسی و بررسی داده ها و اعتبار سنجی کیفیت داده ها پرداخته شده است. مطالعه ی حاضر از نوع توصیفی- مقطعی بوده و مجموعه داده های آن متعلق به بیمارستان های خاتم الانبیاء(ع) و علی ابن ابی طالب(ع) زاهدان می باشد. نمونه گیری به روش سرشماری بوده و ۱۲۵ بیمار را در فاصله ی زمستان ۱۳۹۳ تا تابستان ۱۳۹۴ شامل می گردد. میانگین سن بیماران ۵۰ سال و ۴۲ درصد آن ها مرد و مابقی زن هستند. ۹ درصد بیماران سیگاری، ۸۴ درصد دارای فشارخون، ۸۲ درصد دارای چربی خون و ۶۱ درصد دارای قند خون هستند. ویژگی های آزمایشگاهی بیماران در این مرحله بررسی و شناسایی شد. در مهم ترین گام تحقیق(آماده سازی داده ها یا پیش پردازش داده ها) به بررسی پرونده ی بیماران پرداخته شده است. در جهان واقعی، داده همیشه کامل نیست و در مورد اطلاعات پزشکی این موضوع همیشه درست است. برای حذف تعدادی از تناقض ها و داده های ناقص در ارتباط با داده ها از پردازش داده استفاده گردید. بسیاری از تکنیک های پردازش داده توسط عامری و همکاران ارایه شده است(۵).

در مرحله ی آماده سازی داده ها جهت پاکسازی مجموعه داده با نظر افراد خبره اعمال زیر انجام گرفت: ابتدا فیلدهای خالی برخی از پرونده های بیماران قلبی با مراجعه به نرم افزار پردازش گر و بررسی سابقه پزشکی افراد از قبیل سابقه ی داروهای مصرفی و آزمایش های انجام شده، مقداردهی گردید. رکوردهایی که در این مرحله همچنان دارای مقادیر از دست رفته (Missing Values) بودند، حذف گردیدند. همچنین برخی از رکوردها دارای داده‌های پرت(Outlier) بودند که بر اساس میانگین گیری و استنتاج از داده های هم نوع بهینه شدند و برخی رکوردها که هیچ وجه تشابهی با سایر داده ها نداشتند، حذف گردیدند. لذا در مرحله آماده سازی داده ها تعداد بیماران مورد بررسی به ۱۱۸ مورد رسید. داده ها با مراجعه ی مستقیم پژوهش گر به صورت فایل اکسل تهیه گردید و محتوای داده ها مورد تایید متخصصان حوزه ی مربوط می باشد. در نهایت ۱۲۵ رکورد اولیه از بیماران پس از پالایش و حذف برخی رکوردها به ۱۱۸ رکورد نهایی کاهش یافت.

جدول ۱: داده ها و نوع آن ها پس از پاکسازی

مشخصه	توضیحات	نوع
Sex	جنسیت	اسمی
Age	سن	عددی
Blood Pressure	فشارخون	رتبه ای
Blood Sugar	قند خون	رتبه ای
Cholesterol	کلسترول	رتبه ای
Cigarette	سیگار	رتبه ای
Weight	وزن	رتبه ای
Job	شغل	رتبه ای
Cure Method	روش درمان	اسمی

[Downloaded from journals.tums.ac.ir on 2024-11-17]

هر رکورد دارای ۹ صفت می باشد که مشخصات صفات شامل نوع آن ها در جدول ۱ داده شده است.

ج) مدل سازی

روش های داده کاوی بسیاری برای مدل سازی وجود دارد. در این فاز با استفاده از تکنیک های مختلف داده کاوی به تولید مدل و الگوی بهینه پرداخته شده است. دسته بندی به عنوان یکی از شناخته شده ترین روش های داده کاوی از دو مرحله تشکیل می شود. در مرحله ی اول که مرحله ی استنتاج است هدف کشف مدلی برای تعریف دسته هایی از پیش مشخص شده ی داده هاست. مدل بر اساس نمونه های آموزشی ارایه شده به سیستم ایجاد می شود. الگوریتم استنتاج با استفاده از مقادیر مشخصه های نمونه هایی که به هر دسته تعلق دارند، تعریفی برای آن دسته خاص ایجاد می کند. در مرحله دوم که پیش بینی نام دارد، برای نمونه هایی که تعلق آن ها به دسته خاصی مشخص نیست، بر اساس مدل استنتاج شده می توان تعلق آن ها را پیش بینی نمود(۱۶).

درخت تصمیم برای دسته بندی مورد استفاده قرار می گیرد. معمولاً مجموعه های قوانین مهمترین اطلاعات مرتبط با درخت تصمیم را در بردارند. درخت بر اساس یک معیار تقسیم شاخه به زیر شاخه هایی شکسته می شود و این روند به صورت تناوبی ادامه می یابد تا در نهایت، داده های هر گروه در یک دسته قرار داشته باشند(۱۷). برای بیان قوانین استخراج شده، مسیر ریشه تا برگ درخت پیمایش می شود و قوانین به صورت شرطی بیان می شود.

در پژوهش حاضر مدل سازی با استفاده از نرم افزار SPSS Clementine 12.0 انجام شده است. روش کار داده کاوی پیش بینانه می باشد و از الگوریتم های درخت تصمیم استفاده شده است تا بهترین نسبت بین فیلدهای مختلف به دست آید(۱۷). اجرای الگوریتم های درخت تصمیم شامل C5.0، C&R Tree، CHAID و QUEST و شبکه عصبی روی داده های موجود با هدف پیش بینی روش درمان بیماران قلبی صورت گرفت که تحقق این امر با اندازه گیری پارامترهای فشارخون، قند خون و سیگاری بودن به عنوان عوامل مهم و تاثیرگذار در بیماری قلبی صورت گرفت. برای تولید مدل، ابتدا داده های مورد بررسی به دو بخش آموزش و آزمایش تقسیم گردید. داده های بخش آموزش(۸۵ درصد) درخت را تولید می نمایند و داده های بخش آزمایش(۱۵ درصد) درخت تولید شده را تست و برچسب مربوط به رکوردهای مذکور را تعیین می نمایند. برای آموزش درخت تصمیم یک متغیر طبقه ای باید فیلد خروجی باشد و یک یا تعداد بیشتری فیلد ورودی وجود داشته باشد. فیلدهای ورودی مقادیر به دست آمده از آزمایش های بیماران و خروجی نوع درمان در نظر گرفته شده برای آن ها بود.

جدول ۲: برچسب دسته مدل

برچسب دسته	توضیحات
۱	آنژیوگرافی: در این روش تعداد عروق کرونر مسدود شده، محل انسداد و میزان آن مشخص می‌شود و مستقیم ترین راه کشف مشکلات شریان‌های کرونری قلب است.
۲	آنژیوپلاستی: دراین روش سرخرگ های کرونر تنگ یا مسدود شده به وسیله ی رسوب چربی و لخته، بدون نیاز به عمل جراحی باز می گردند.
۳	درمان طبی: در این روش با آگاهی دادن به بیمار و تجویز داروهای مربوط به بیماری به بهبود سلامت بیمار کمک می شود.
٤	Pace maker: به بخشی از قلب و یا در حالت مصنوعی دستگاهی که عمل ضربان سازی را تقلید می‌کند گفته می‌شود که ضربان ایجاد کرده و آهنگ آن را تنظیم می‌کند.
٥	قلب باز: در صورتی که تعداد عروق تنگ، زیاد بوده و یا تنگی در محل خطرناک نظیر تنه اصلی سرخرگ چپ یا در ابتدای شریان کرونری قلب یا محل دوشاخه شدن عروق کرونر باشد، امکان آنژیوپلاستی وجود نداشته، نیاز به عمل جراحی پیوند عروق قلب می باشد.

جدول ۲ برچسب دسته(روش درمان) را به همراه توضیحات نشان می دهد.

د) ارزیابی

پس از اجرای مدل سازی باید به ارزیابی نتایج حاصل پرداخت. نتایج ارزیابی باعث بهبود مدل می شود و مدل را قابل استفاده می نماید. شاخص های مختلفی مانند شفافیت(Specificity)، حساسیت(Sensitivity)، دقت(Precision) و صحت(Accuracy) برای ارزیابی روش های دسته بندی وجود دارند که طبق روابط ۱ تا ۴ محاسبه می گردند. برای محاسبه ی میزان شاخص ها می توان از ماتریس اغتشاش(Confusion Matrix) استفاده کرد. این ماتریس ابزار مفیدی برای تحلیل چگونگی عملکرد روش دسته بندی در تشخیص داده ها یا مشاهدات دسته های مختلف است. حالت

[Downloaded from journals.tums.ac.ir on 2024-11-17]

ایده آل این است که بیشتر داده های مرتبط با مشاهدات روی قطر اصلی ماتریس قرار گرفته باشند و مابقی مقادیر ماتریس صفر یا نزدیک صفر باشند(۱۸و۱۴).

TN: بیانگر تعداد رکوردهایی است که دسته ی واقعی آنها منفی بوده و الگوریتم دسته بندی نیز دسته ی آنها را به درستی منفی تشخیص داده است.

TP: بیانگر تعداد رکوردهایی است که دسته ی واقعی آنها مثبت بوده و الگوریتم دسته بندی نیز دسته ی آنها را به درستی مثبت تشخیص داده است.

FP: بیانگر تعداد رکوردهایی است که دسته ی واقعی آنها منفی بوده و الگوریتم دسته بندی دسته ی آنها را به اشتباه مثبت تشخیص داده است.

FN: بیانگر تعداد رکوردهایی است که دسته ی واقعی آنها مثبت بوده و الگوریتم دسته بندی دسته ی آنها را به اشتباه منفی تشخیص داده است.

$$Specificity = \frac{TN}{FP + TN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Percision = \frac{TP}{TP + FP} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

و) توسعه

ساخت مدل، پایان یک پروژه نیست و هدف از پروژه های داده کاوی کشف دانش و استفاده از دانش کشف شده به صورت عملی در آینده می باشد. در حقیقت هدف از انجام مراحل مختلف کشف دانش، دست یابی به نتایجی است که بتوان از آن ها در دنیای واقعی و برای بهبود کارایی سازمان ها استفاده کرد. دانش کشف شده باید سازماندهی شود و به شکل قابل استفاده برای دیگران نیز در آید. به طور کلی وجود رویکردی سازمان دهی شده جهت پیش بینی روش درمان بیماران قلبی به منظور کمک به پزشک برای افزایش صحت تشخیص و جلوگیری از عوارض ناشی از تشخیص نادرست ضروری است. نقطه ی قابل بهبود در حیطه ی مورد بررسی، ایجاد سیستمی پزشک یار جهت کمک به پزشکان در تشخیص روش درمان می باشد.

یافته ها

الگوریتم های C5.0، C&R Tree، CHAID و QUEST و شبکه عصبی مصنوعی روی مجموعه داده ی مورد بررسی اجرا گردید.

جدول ۳: مقدار صحت برای مدل های تولید شده

نوع الگوریتم	نام الگوریتم	داده آموزش(درصد)	داده آزمایش(درصد)
درخت تصمیم	C&R Tree	۷۶/۰۴	۵۶/۵۲
	QUEST	۶۶/۶۷	۵۶/۵۲
	CHAID	۶۳/۵۴	۵۲/۱۷
	C5.0	۶۵/۶۲	۴۷/۸۳
شبکه عصبی	Neural Net	۵۷/۲۹	۳۹/۱۳

مقدار صحت مدل های تولید شده برای داده های آموزش و آزمایش طبق جدول ۳ می باشد. بیشترین میزان صحت با استفاده از گره درخت C&R Tree به دست آمد. بنابراین از درخت C&R Tree برای تعیین روش درمان استفاده شده است.

جدول ۴: مقدار شاخص ها برای الگوریتم C&R Tree

برچسب دسته	شفافیت	حساسیت	دقت	صحت
۱	٪۳۴	٪۷۱	٪۹۲	٪۷۶
۲	٪۹۷	٪۱۰۰	٪۴۳	٪۹۷
۳	٪۸۹	٪۹۲	٪۴۲	٪۸۷
٤	٪۹۷	٪۶۰	٪۶۰	٪۹۷
٥	٪۹۱	٪۶۳	٪۵۹	٪۸۹

[Downloaded from journals.tums.ac.ir on 2024-11-17]

جدول ۴ مقادیر چهار شاخص برای هر کدام از برچسب دسته ها را نشان می دهد که برای الگوریتم C&R Tree و با استفاده از ماتریس اغتشاش محاسبه گردیده است. مقدار ۸۹/۲ درصد برای میانگین کلی صحت مدل، ۵۹/۲ درصد برای دقت، ۸۱/۶ درصد برای شفافیت و ۷۷/۲ درصد برای حساسیت توسط الگوریتم C&R Tree نشان می دهد که درخت تولیده شده می تواند قوانین جامعی برای پیش بینی وضعیت بیماران آینده ارایه نماید.

جدول ۵: قواعد تلازمی تولید شده توسط درخت تصمیم C&R Tree

برچسب دسته	قوانین
۱	if Cholesterol <= 0.5 and Age<= 66 and Blood Pressure <= 0.5 and Age <= 32.5 and Age > 29.5 Then Label=1
	if Cholesterol > 0.5 and Cigarette <= 0.5 and Blood Sugar <= 0.5 and Blood Pressure > 0.5 Then Label=1
	if Cholesterol > 0.5 and Cigarette <= 0.5 and Blood Sugar > 0.5 and Age > 37 Then Label=1
	if Cholesterol > 0.5 and Cigarette > 0.5 and Blood Pressure > 0.5 and Blood Sugar <= 0.5 Then Label=1
۲	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure > 0.5 and Sex=M and Age <= 53.5 Then Label=2
	if Cholesterol <= 0.5 and Age > 66 and Age > 70.5 and Sex=F and Blood Sugar > 0.5 Then Label=2
۳	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure <= 0.5 and Age <= 32.5 and Age <= 29.5 Then Label=3
	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure <= 0.5 and Age > 32.5 Then Label=3
	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure > 0.5 and Sex=F and Age <= 53.5 Then Label=3
	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure > 0.5 and Sex=M and Age > 53.5 Then Label=3
	if Cholesterol > 0.5 and Cigarette <= 0.5 and Blood Sugar <= 0.5 and Blood Pressure <= 0.5 and Age <= 67.5 Then Label=3
۴	if Cholesterol <= 0.5 and Age > 66 and Age > 70.5 and Sex=F and Blood Sugar <= 0.500 Then Label=4
	if Cholesterol <= 0.5 and Age > 66 and Age > 70.5 and Sex=M Then Label=4
	if Cholesterol > 0.5 and Cigarette <= 0.5 and Blood Sugar <= 0.5 and Blood Pressure <= 0.5 and Age > 67.5 Then Label=4
۵	if Cholesterol <= 0.5 and Age <= 66 and Blood Pressure > 0.5 and Sex=F and Age > 53.5 Then Label=5
	if Cholesterol <= 0.5 and Age > 66 and Age <= 70.5 Then Label=5
	if Cholesterol > 0.5 and Cigarette <= 0.5 and Blood Sugar > 0.5 and Age <= 37 Then Label=5
	if Cholesterol > 0.5 and Cigarette > 0.5 and Blood Pressure <= 0.5 Then Label=5
	if Cholesterol > 0.5 and Cigarette > 0.5 and Blood Pressure > 0.5 and Blood Sugar > 0.5 Then Label=5

دانش یا قوانین استخراج شده در جدول ۵ نشان داده شده است.

بحث

از داده کاوی برای به دست آوردن روابط مفید بین عوامل خطر زا در بیماری های قلب و عروق استفاده می شود. این بیماری ها با توجه به شیوع و سهمی که در مرگ و میر انسان ها دارند از اهمیت بالایی برخوردارند. ویژگی متمایز مطالعه ی حاضر کشف الگویی برای شناسایی روش درمان بیماران قلبی با استفاده از الگوریتم های درخت تصمیم و شبکه عصبی است. از بین الگوریتم های مورد استفاده، بهترین نتایج از الگوریتم C&R Tree با دقت ۵۹/۲ و صحت ۷۶/۰۴ به دست آمد. تاثیرگذارترین پارامترها در انتخاب روش درمان بیماری قلبی کلسترول، سیگاری بودن، قند خون، فشارخون و سن شناخته شده اند. پارامتر جنسیت اهمیت کمتری نسبت به سایر پارامترها دارد. با استفاده از قوانین ایجاد شده برای یک نمونه ی جدید با ویژگی های مشخص، می توان پیش بینی کرد که چه روش درمانی مناسب نمونه ی جدید است. به طور کلی دانش استخراج شده از مدل های داده کاوی می تواند به عنوان یک سیستم تصمیم یار در دنیای واقعی استفاده شود.

جدول ۵ نشان می دهد که اگر فرد کلسترول و فشارخون نداشته باشد و سن وی بین ۲۹/۵ و ۳۲/۵ باشد برچسب دسته یک می شود و روش آنژیوگرافی برای درمان وی مناسب است. دانش استخراج شده از مدل C&R Tree می تواند به عنوان الگویی جهت پیش بینی روش درمانی مناسب توصیه شود که این قوانین پارامترهای اثرگذار را در روش درمانی مشخص می کنند. قواعد تلازمی ارایه شده

[Downloaded from journals.tums.ac.ir on 2024-11-17]

در جدول ۵ نحوه ی پیش بینی مدل را نشان می دهد. هر چه تعداد قوانین تولید شده بیشتر باشد بهتر است؛ زیرا این نشان می دهد که مدل پیش بینی کننده، جزئیات بیشتری را مد نظر قرار داده است.

مقادیر بالا برای شاخص های حساسیت، شفافیت، صحت و دقت نشان دهنده ی این است که طبقه بندی مورد استفاده نمونه های بیشتری را در جای درست خود طبقه بندی کرده است. Ashwinkumar و Anandakumar محدودیت های اجتماعی، قانونی و اخلاقی روی داده های پزشکی از جمله بیماری قلبی را بررسی کرده اند(۱۹). Palaniappan و Awang، Venkatalakshmi و Shivsankar و Srinivas و همکاران در مطالعات خارجی به ارایه سیستم هایی برای پیش بینی بیماری قلبی پرداخته اند(۹و۷و۶). Guru و همکاران پیش بینی بیماری قلبی را بر اساس پارامترهای فشار خون، قند، شکر و غیره انجام داد. جهت انجام کار ۷۸ پرونده با ۱۳ ویژگی به عنوان مجموعه آموزش و آزمایش استفاده شد(۲۰). Parthiban و Subramanian داده کاوی شبکه عصبی را با قابلیت های یادگیری با منطق فازی که دارای رویکردی کیفی است ترکیب کرده و از آن برای تشخیص بیماری قلبی استفاده کرده اند(۲۱).

Masethe و Masethe، Soni و همکاران و Tanejaaneja از داده کاوی به منظور پیش بینی حمله ی قلبی استفاده نمودند(۱۱و۱۰و۸). در مطالعات داخلی نیز رحیمی شاطرانلو و علیزاده پیش بینی بیماری کرونری قلبی، صفدری و همکاران پیش بینی ابتلا به آنفارکتوس قلبی، علیزاده ثانی تشخیص بیماری های قلبی و محمودی و همکاران پیش بینی عروق کرونر را با استفاده از الگوریتم های داده کاوی انجام دادند(۱۵-۱۲). شفیعی و ابراهیمی با بهره گیری از الگوریتم های داده کاوی مدلی جهت تشخیص بیماری عروق کرونری معرفی نمودند و نشان دادند که متغییر اسکن-تالیوم مهم ترین ویژگی در تشخیص بیماری های قلبی می باشد(۲۲).

بررسی مطالعات موجود نشان می دهد که تاکنون از داده کاوی برای تشخیص بیماری قلبی و پارامترهای تاثیرگذار در ایجاد بیماری قلبی استفاده گردیده است. از نظر هدف هیچ یک از مطالعات داخلی و خارجی انجام شده تاکنون به پیش بینی روش درمان بیماری قلبی نپرداخته اند و اکثر مطالعات روی تشخیص بیماری تمرکز داشته اند. در حالی که پژوهش حاضر مدلی برای پیش بینی روش درمان بیماری قلبی ارایه داده است. از نظر روش مدل سازی، در اکثر مطالعات انجام شده مدل تولید شده با الگوریتم های درخت تصمیم بر سایر مدل ها چون شبکه عصبی برتری داشته اند(۱۳-۹و۷)، که در این مطالعه هم این نتیجه تایید گردید. عدم ثبت صفاتی چون درصد انسداد رگ و منبع انتخاب روش درمان می تواند جزو محدودیت های سیستم ارایه شده محسوب گردد. همچنین محدود بودن تعداد نمونه های با روش درمان Pace maker را می توان جزو محدودیت های سیستم به شمار آورد.

نتیجه گیری

داده کاوی روی داده های پزشکی از اهمیت بالایی برخوردار است و طراحی سیستم های تصمیم یار جهت یاری رساندن به پزشکان در زمینه ی تشخیص نوع بیماری یا انتخاب نوع درمان مناسب، با کمک داده کاوی می تواند کمک شایانی در زمینه ی نجات جان انسان ها انجام دهد. در همین راستا در پژوهش حاضر الگوریتم C&R Tree با بهترین عملکرد به پیش بینی روش درمان مناسب بیماران قلبی پرداخته است. طبق نظر متخصصان قلب و عروق می توان گفت که ریسک پارامترهای سن بالا، مصرف سیگار، فشارخون بالا، چربی خون بالا بیشترین تاثیر را در روش درمانی لحاظ شده دارند و این در حالی است که اولویت بندی متغیرها توسط الگوریتم C&R Tree، این متغیرها را جزو پارامترهای تاثیرگذار قرار داده است که نشان از اهمیت این متغیرها در روش درمانی انجام شده دارد.

همچنین نتایج حاصل شده نشان می دهد که صحت پیش بینی درخت تصمیم روی داده های مورد بررسی از شبکه ی عصبی که یک روش پرکاربرد و مشهور می باشد، بیشتر است. استفاده از صفات درصد انسداد رگ و منبع انتخاب روش درمان به همراه صفات موجود می تواند سیستمی جامع جهت پیش بینی روش درمانی مناسب برای بیماران قلبی ارایه نماید و بدین طریق با کاهش هزینه های درمان بیماران و افزایش کیفیت ارایه خدمات به آن ها می توان گامی موثر در جهت ارتقای سیستماتیک تشخیص پزشکی برداشت.

تشکر و قدردانی

بدین وسیله از کلیه پرسنل محترم بیمارستان های خاتم الانبیاء(ع) و علی ابن ابی طالب(ع) زاهدان تقدیر و تشکر می شود. پژوهش حاضر حاصل یک طرح تحقیقاتی با شماره ۳۳ مصوب سازمان بیمه سلامت ایران و بدون حمایت مالی می باشد.

منابع

1. Chaitrali SD & Sulabha SA. A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology 2012; 3(3): 30-40.

[Downloaded from journals.tums.ac.ir on 2024-11-17]

2. Karimi S, Javadi M & Jafarzadeh F. Economic burden and costs of chronic diseases in Iran and the world. Health Information Management 2012; 8(7): 984-96[Article in Persian].

3. Jooriyan N & Ashoori M. Predicting the effectiveness of preeclampsia medications based on dose and method of drug consumption using data mining. Iranian Journal of Obstetrics, Gynecology and Infertility 2014; 17(123): 13-22[Article in Persian].

4. Alizadeh S, Ghazanfari M & Teimorpour B. Data mining and knowledge discovery. 2nd ed. Tehran: Publication of Iran University of Science and Technology; 2011: 70-250[Book in Persian].

5. Ameri H, Alizadeh S & Barzegari A. Knowledge extraction of diabetics' data by decision tree method. Health Management 2013; 16(53): 58-72[Article in Persian].

6. Palaniappan S & Awang R. Intelligent heart disease prediction system using data mining techniques. International Journal of Computer Science and Network Security 2008; 8(8): 343-50.

7. Venkatalakshmi B & Shivsankar MV. Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology 2014; 3(3): 1873-7.

8. Masethe HD & Masethe MA. Prediction of heart disease using classification algorithms, USA: Proceedings of The World Congress on Engineering and Computer Science, 2014.

9. Srinivas B, Kavihta KA, Govrdhan R & Karimnagar J. Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering 2010; 2(1): 250-5.

10. 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.

11. Tanejaaneja A. Heart disease prediction system using data mining techniques. Journal of Computer Science & Technology 2013; 6(4): 457-66.

12. Rahimi Shateranloo E & Alizadeh S. Prediction of coronary heart disease using hybrid data mining. Journal of Soft Computing and information Technology 2014; 3(1): 56-62[Article in Persian].

13. 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[Article in Persian].

14. Alizadeh Sani R. Diagnosis of heart disease using data mining [Thesis in Persian]. Tehran: Sharif University of Technology; 2012.

15. Mahmoudi I, Askari Moghadam R, Moazzam M & Sadeghian S. Prediction model for coronary artery disease using neural networks and feature selection based on classification and regression tree. Journal of Shahrekord University of Medical Sciences 2013; 15(5): 47-56[Article in Persian].

16. Ashoori M, Naji Moghaddam V, Alizadeh S & Safi M. Classification and clustering algorithm application for prediction of tablet numbers: Case study diabetes disease. Health Information Management 2013; 10(5): 739-49[Article in Persian].

17. Chen G & Astebro T. How to deal with missing categorical data: Test of a simple bayesian method. Organizational Research Methods 2003; 6(3): 309-27.

18. Han J & Kamber M. Data mining: Concepts and techniques. 2nd ed. United States: Morgan Kaufman; 2006: 360-2.

19. Ashwinkumar UM & Anandakumar KR. Ethical and legal issues for medical data mining. International Journal of Computer Applications 2010; 1(28): 7-11.

20. Guru N, Dahiya A & Rajpal N. Decision support system for heart disease diagnosis using neural network. Delhi Business Review 2007; 8(1): 1-6.

21. Parthiban L & Subramanian R. Intelligent heart disease prediction system using canfis and genetic algorithm. International Journal of Biological & Medical Sciences 2008; 3(3): 157-64.

22. Shafiee H & Ebrahimi M. Accurate prediction of coronary artery disease using bioinformatics algorithms. Qom Univ Med Sci 2016; 10(4): 22-35[Article in Persian].

[Downloaded from journals.tums.ac.ir on 2024-11-17]

A Model to Predict Heart Disease Treatment Using Data Mining

Mazaheri Sajad[1] (M.S.) – Ashoori Maryam[2] (M.S.) – Bechari Zeynab[3] (B.S.)

1 Master of Science in Computer Engineering Software, The Health Insurance Office of Sistan & Balouchestan Province, Iran Health Insurance Organization, Zahedan, Iran

2 Instructor, Information Technology Department, Higher Educational Complex of Saravan, Saravan, Iran

3 Bachelor of Science in Nursing, The Health Insurance Office of Sistan & Balouchestan Province, Iran Health Insurance Organization, Zahedan, Iran

Abstract

Received: Nov 2016
Accepted: Mar 2017

Background and Aim: Nowadays heart disease is very common and is a major cause of mortality. Proper and early diagnosis of this disease is very important. Diagnostic methods and treatments of the disease are so expensive and have many side effects. Therefore, researchers are looking for cheaper ways to diagnose it with high precision. This study aimed to identify a model for the treatment of heart disease.

Materials and Methods: In this descriptive cross-sectional study, the sampling method was census. The sample consisted of data from Khatam and Ali Ibn Abi Talib Hospitals in Zahedan. The data were developed as an Excel file, and Clementine12.0 software was used for data analysis. In the present study, C5.0, C & R Tree, CHAID, and QUEST algorithms and artificial neural network were carried out on the collected data.

Results: The accuracy of 76.04 by C & R algorithm indicates the better performance of Decision Tree Algorithms than that of the Neural Network.

Conclusion: This study aimed to provide a model for the prediction of a suitable heart disease treatment to reduce treatment costs and provide better quality of services for physicians. Due to considerable implementation risks of invasive diagnostic procedures such as angiography and also obtaining successful experiences of data analysis in medicine, this study has presented a model based on data analysis techniques. The improvable point of this model is the provision of a decision support system to help physicians to increase the accuracy of diagnosis in the treatment of diseases.

Keywords: Data Mining, Heart Disease, Prediction, Cure Method, Decision Tree, Neural Network

*** Corresponding Author:**
Mazaheri S;
Email:
sajad.mazaheri@gmail.com

[Downloaded from journals.tums.ac.ir on 2024-11-17]