داده‌کاوی و کاربرد آن در پزشکی

نویسنده: وحید عاشوری-ashoori.vahid2015@gmail.com

چکیده: امروزه در دانش پزشکی شاهد جمع‌آوری داده‌های فراوان در مورد بیمارهای مختلف هستیم. مراکز پزشکی با مقاصد گوناگون به جمع‌آوری این داده‌ها می‌پردازند. تحقیق روی این داده‌ها و به دست آوردن نتایج و الگوهای مفید در رابطه با بیماری‌ها یکی از اهداف استفاده از این داده‌ها می‌باشد. حجم بالای این داده‌ها و سردرگمی حاصل از آن مشکلی است که مانع رسیدن به نتایج قابل‌قبول می‌شود. داده‌کاوی[1] می‌تواند اطلاعات ارزشمندی را از پایگاه داده‌های پزشکی جهت کشف دانش استخراج نماید. هدف این مقاله، که از نوع مروری می‌باشد، این است که مفاهیم مرتبط با داده‌کاوی و کاربرد آن در پزشکی را مطرح کند.

کلیدواژه: Data Mining، پایگاه داده‌های پزشکی، داده‌کاوی در پزشکی، کشف دانش، سیستم‌های کمک تصمیم.

1- مقدمه

بدون تردید تمدن بشری مدیون پیشرفت امکانات ثبت و ذخیره-سازی و پردازش داده‌ها می‌باشد که به همراه خود دیگر زمینه-های تمدن بشری را به این حد تکامل رسانده است. داده‌ها نمایشی از واقعیت‌ها، معلومات، مفاهیم، رویدادها یا پدیده‌ها برای برقراری ارتباط، تفسیر یا پردازش، توسط انسان یا ماشین می-باشند[26].

پیشرفت فوق‌العاده در کسب و ذخیره‌سازی داده‌های عددی، سبب به وجود آمدن پایگاه داده‌های بزرگ‌شده است. داده‌های تبادلات تجاری، کشاورزی، ترافیک، اینترنت، داده‌های نجومی، جزئیات مکالمات تلفنی، داده‌های پزشکی و درمانگاهی مثال‌هایی از چنین پایگاه داده‌هایی می‌باشند. درواقع تکنیک‌های تولید و جمع‌آوری پایگاه داده‌ها بسیار سریع‌تر از توانایی ما در درک و استفاده از آن-ها رشد کرده است[26].

امروزه افزایش سریع حجم پایگاه داده‌ها به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای پرقدرت میسر نیست. در طول دهه‌ی گذشته با پیشرفت روزافزون پایگاه داده‌ها، حجم داده‌های ثبت‌شده به‌طور متوسط هر 5 سال 2 برابر می‌شود. در این میان سازمان‌هایی موفق هستند که بتوانند حداقل 7٪ داده-هایشان را تحلیل کنند. تحقیقات انجام‌شده نشان می‌دهد که سازمان‌ها کمتر از یک درصد داده‌هایشان را برای تحلیل استفاده می‌کنند. به‌عبارت‌دیگر درحالی‌که غرق در داده‌ها هستند، تشنه‌ی دانش می‌باشند[27].

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره‌شده در این سیستم‌ها، به ابزاری نیاز است تا بتوان این داده‌ها را پردازش کرد و از داده‌کاوی، مجموعه‌ای از تکنیک‌ها است که توسط آن‌ها، دانش پنهان درون داده‌ها و الگوهای ناشناخته آن‌ها کشف می‌شود. درواقع فرآیند انتخاب، کاوش و مدل بندی داده‌های حجیم، جهت کشف روابط نهفته باهدف به دست آوردن نتایج واضح و مفید را داده‌کاوی گویند[30].

امروزه حجم داده‌های پزشکی که به‌صورت الکترونیکی ذخیره می‌شوند روزبه‌روز در حال افزایش است اما مجموعه‌ی عظیم داده‌های خام به‌خودی‌خود کاربردی ندارد. برای معنی بخشیدن به این داده‌ها باید آن‌ها را تحلیل و تبدیل به اطلاعات نموده و یا بهتر از آن، به دانش تبدیل کرد.

از طرف دیگر در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف از اهمیت فراوانی برخوردار است. تحقیق روی این داده‌ها و به دست آوردن نتایج و الگوهای مفید روی بیماری‌های مختلف، یکی از اهداف استفاده از این داده‌ها است. با به‌کارگیری الگوریتم‌های داده‌کاوی می‌توان اطلاعات و دانش مفیدی را از این داده‌ها اکتشاف نمود و پزشکان را در قضاوت صحیح یاری رسانید.

بنابراین به‌کارگیری داده‌کاوی بر روی‌داده‌های پزشکی، دستاوردهای حیاتی، باارزش و اثرگذاری را به ارمغان آورده و ازجمله می‌تواند برای یاری‌رسانی به پزشکان در زمینه‌های مختلف ازجمله تجزیه‌وتحلیل داده‌های موجود در سیستم‌های اطلاعات سلامت، تعیین و پیش‌بینی میزان موفقیت اقدامات پزشکی، تشخیص و پیش‌بینی انواع بیماری‌ها، تعیین روش درمان بیماری-ها که منجر به نجات عده زیادی از جان بیماران می‌شود، باشد.

در بخش 2 به معرفی داده‌کاوی و دلایل استفاده از آن و بخش 3 نیز به بررسی داده‌کاوی درزمینه‌ی پزشکی پرداخته شد. قسمت 4

روش‌ها و تکنیک‌های داده‌کاوی معرفی شدند و بخش 5 نمونه-هایی از داده‌کاوی در پزشکی بررسی گردید و در پایان نیز نتیجه-گیری بیان شد.

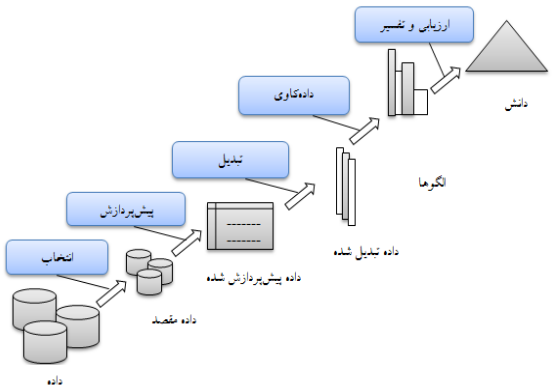
2- داده‌کاوی و دلایل استفاده از آن

داده‌کاوی عبارت است از فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل‌فهم و قابل‌اعتماد از پایگاه داده‌های بزرگ و استفاده از آن در تصمیم‌گیری در فعالیت‌های تجاری مهم.
کشف دانش، فرآیند تشخیص الگوهای معتبر جدید، مفید و قابل‌درک در داده‌ها است[1].
داده‌کاوی مفهومی است که با پایگاه‌های داده‌های بزرگی مانند مخزن داده‌ها و انبار داده‌ها عجین شده است، و به دنبال استخراج اطلاعات مفید و ناشناخته از داده‌های خام است.
درواقع داده‌کاوی شکل پیشرفته‌ی پشتیبانی از تصمیم است و برخلاف ابزارهای پرس‌وجوی غیرفعال بدون الزام به طرح سؤال از طرف کاربر به تولید الگو، روندها و قواعد برنامه‌ریزی‌شده می-پردازد[2]، به‌عبارت‌دیگر قدرت داده‌کاوی در این است که می‌تواند الگوهایی را که در کاربر موردتوجه قرار نگرفته است، افشا کند و پاسخ‌هایی را برای سؤالاتی که هرگز درخواست نشده بود، تولید نماید.

داده‌کاوی یک‌رشته‌ی جدید با کاربردهای وسیع و گوناگون است که بنا به اعلام دانشگاه MIT یکی از 10 دانش برتر است که منجر به ایجاد تحول در عصر تکنولوژی می‌شود و در تمام زمینه-ها کاربرد دارد. اصولاً هرجایی که داده وجود داشته باشد داده-کاوی نیز معنا می‌یابد[3].
همان‌طور در شکل 1 ملاحظه می‌شود مراحل داده‌کاوی به شرح ذیل می‌باشد:

- انتخاب داده‌ها: برای کاهش هزینه‌های عملیات داده-کاوی، داده‌هایی از پایگاه داده انتخاب می‌شوند که موردمطالعه هستند.
- پیش‌پردازش داده‌ها: پاک‌سازی داده‌ها یعنی از بین بردن نویز و ناسازگاری آن‌ها و یکپارچه‌سازی داده‌ها یعنی ترکیب چندین منبع داده صورت می‌پذیرد.
- تبدیل داده‌ها: برای انجام عملیات داده‌کاوی لزوماً باید تبدیلات خاصی روی‌داده‌ها انجام گیرد.
- داده‌کاوی: با استفاده از تکنیک‌های داده‌کاوی داده‌ها مورد کاوش قرارگرفته، دانش نهفته در آن‌ها استخراج‌شده و الگوسازی صورت می‌گیرد.
- ارزیابی الگوها و تفسیر نتیجه: نتایج و الگوهای ارائه‌شده توسط ابزار داده کاو، موردبررسی قرارگرفته و نتایج مفید معین می‌شود[26].

لازم به ذکر است که آماده‌سازی داده‌ها، حدود 60 تا 90 درصد زمان موردنیاز برای کاوش داده را صرف کرده و 75 تا 90 درصد موفقیت پروژه‌های داده‌کاوی به آن مربوط می‌شود[4].



شکل 1: مراحل داده‌کاوی

Wickramasinghe و همکارانش مهم‌ترین دلایل استفاده از داده‌کاوی را افزایش حجم داده‌های موجود و آتی سازمان‌ها عنوان می‌کنند که نیاز به پردازشی فراتر از رویکردهای سنتی دارد. برای انسان مواجه‌شدن با این حجم انبوه و در حال رشد داده‌ها با اتکای محض به شیوه‌های سنتی غیرممکن است. دلایل دیگری نیز وجود دارند که بر اهمیت و ضرورت استفاده از داده‌کاوی تأکید می‌کنند که ازجمله می‌توان در دسترس بودن حجم زیادی از داده‌ها، افزایش رقابت جهانی، مهیا بودن تکنولوژی اشاره کرد[5].

3- داده‌کاوی درزمینه‌ی پزشکی

مفهوم کشف دانش از داده‌ها بیش از یک دهه است که در محیط-های مالی و تجاری در حال استفاده می‌باشد. در علوم مدیریت ارتباطات، مهندسی، وب‌کاوی، تحلیل جرائم و پزشکی نیز جای خود را بازکرده است. اگرچه کشف دانش باهدف شناسایی اختلاس مالی وارد عرصه‌ی سلامت و پزشکی شد، اما به‌تدریج در حوزه‌ی بالینی نیز مورداستفاده قرار گرفت. این مهم ناشی از

www.kahrobaonline.ir

تغییر سریع هوشیاری نسبت به اطلاعات در حوزه‌ی سلامت است. حمایت از تصمیم‌گیری‌های پزشکی با استفاده از داده‌های واقعی و شواهد، قرن‌ها وجود داشته است. محققان و پژوهشگران درگذشته به دلیل قابل‌کنترل بودن حجم اطلاعات، شخصاً قادر به جمع-آوری، غربال کردن و تجزیه‌وتحلیل داده‌های مربوط به مرگ‌ومیر در طول زمان بودند. اما امروزه اندازه‌ی جمعیت، حجم داده‌های الکترونیکی جمع‌آوری‌شده همراه با جهانی‌شدن و سرعت شیوع بیماری‌ها تقریباً این امر را غیرممکن می‌سازد. اینجاست که داده-کاوی برای مراقبت‌های پزشکی، مفید و قابل‌اطمینان بوده و با کشف دانش به مقابله با مشکلات مختلف در بخش بهداشت و درمان می‌پردازد.

حوزه‌ی پزشکی از بخش‌های مهم در جوامع صنعتی است. استخراج دانایی از میان حجم انبوه داده‌های مرتبط با سوابق بیماری و پرونده‌های پزشکی افراد با استفاده از فرآیند داده‌کاوی می‌تواند منجر به شناسایی قوانین حاکم بر ایجاد و رشد بیماری‌ها شده و اطلاعات ارزشمندی را به‌منظور شناسایی علل رخداد بیماری‌ها، تشخیص، پیش‌بینی و درمان بیماری‌ها با توجه به عوامل محیطی حاکم در اختیار متخصصان و دست‌اندازان حوزه‌ی پزشکی قرار دهد. نتیجه‌ی این مسئله به معنای افزایش عمر و ایجاد آرامش برای افراد جامعه است که خود می‌تواند اساسی‌ترین عامل مهم جهت استفاده از تکنیک‌های داده‌کاوی در پزشکی باشد.

4- روش‌ها و تکنیک‌های داده‌کاوی

به‌طورکلی هدف داده‌کاوی، یادگیری و آموختن از داده‌ها است و بر این اساس دودسته‌ی کلی از استراتژی‌های داده‌کاوی شامل یادگیری نظارت‌شده و یادگیری بدون نظارت وجود دارد. شیوه-های یادگیری نظارت‌شده زمانی به کار می‌رود که ارزش متغیرهای ورودی (درون دادها) برای ما شناخته‌شده باشد. این استراتژی باهدف پیش‌بینی داده‌ها و کشف اطلاعات به کار می-رود اما در شیوه‌ی یادگیری بدون نظارت، بر روی‌داده‌هایی که ارزش آن‌ها شناخته‌شده نیست تمرکز می‌شود.

روش‌های اصلی داده‌کاوی دودسته‌ی توصیفی و پیشبینانه می-باشند. هدف از توصیف، یافتن الگوهایی در مورد داده‌هاست که برای انسان قابل تفسیر باشد. وظایف پیشبینانه به‌منظور پیش-بینی رفتارهای آینده‌ی آن‌ها استفاده می‌شوند[30]. لازم به ذکر است که روش‌های توصیفی از نوع یادگیری بدون نظارت و روش-های پیش‌بینانه یادگیری با نظارت هستند.

از تکنیک‌های داده‌کاوی مانند درخت تصمیم‌گیری، شبکه عصبی، خوشه‌بندی، الگوریتم ژنتیک و غیره را می‌توان نام بـرد کـه ادامـه به برخی از آن‌ها اشاره می‌شود.

4-1- شبکه عصبی

شبکه‌ی عصبی روشی است که قصد دارد با استفاده از مدل‌های ریاضی و توان کامپیوتر برخی از جنبه‌های ساده مغز انسان را شبیه‌سازی کند. شبکه‌های عصبی به‌صورت یکی از بخش‌های پیچیده‌ی مغز انسان، به‌عنوان یک ساختار یادگیری غیرقابل‌درک، مشهور شده است. شبکه‌های عصبی مصنوعی در برخی عملیات مانند پیش‌بینی و دسته‌بندی کاربرد دارد[8]. به‌عنوان‌مثال استفاده از شبکه عصبی در تشخیص بیماری که مهم‌ترین کار مرحله درمان است، برای پزشکان بسیار کارا و مفید خواهد بود.

4-2- درخت تصمیم‌گیری

درختان تصمیم ابزار قدرتمند و درعین‌حال رایجی هم برای دسته‌بندی و هم برای پیش‌بینی هستند. در ساختار درخت تصمیم، پیش‌بینی به‌دست‌آمده از درخت در قالب یک‌سری قواعد توضیح داده می‌شود. از ویژگی‌های درخت تصمیم سهولت تفسیر آن می‌باشد. مانند طراحی یک سیستم کمک تصمیمی برای تعیین نوع درمان بیماران مبتلا به سنگ حالب با کمک درخت تصمیم[19].

4-3- الگوریتم ژنتیک

الگوریتم ژنتیک تکنیک‌های بهینه‌ای است برای ارتقای سایر الگوریتم‌های داده‌کاوی، به‌گونه‌ای که از بهترین مدل بر روی مجموعه‌ی داده‌ها استفاده کنند و می‌تواند برای یک بیماری خاص بهترین برنامه‌ی درمانی را تعیین کند[6].

در جدول زیر روش‌ها و الگوریتم‌های داده‌کاوی به کار گرفته‌شده در حوزه پزشکی موردبررسی قرارگرفته است.

جدول 1: دسته‌بندی روش‌ها و الگوریتم‌های داده‌کاوی در علوم پزشکی[18،17،16،15،14و19].

ردیف	حوزه پزشکی	روش داده کاوی	الگوریتم داده کاوی	توضیحات
1	پیشگویی بیماری	خوشه‌بندی دسته‌بندی قوانین وابستگی	K-Means Apriori	تعیین عوامل مؤثر بر ایجاد انواع سرطان
2	تعیین بهترین نوع درمان	خوشه‌بندی دسته‌بندی قوانین وابستگی	K-Means Apriori	تعیین بهترین نوع درمان سرطان
3	بررسی داده های آزمایشگاهی	خوشه‌بندی	SAMBA	بررسی رفتار ژن‌ها با رشته‌هایDNA جهت پیشگویی اختلالات ژنتیکی
4	پیشگویی در بیماران اورژانس	خوشه‌بندی	ترکیب K-Means و الگوریتم SOM	اتخاذ تصمیم درمانی صحیح و به موقع و کاهش هزینه‌های بیمارستانی
5	پیش‌بینی مدت‌زمان بستری	خوشه‌بندی دسته‌بندی	درخت تصمیم شبکه‌های عصبی Naïve Bays C 4.5	پیش‌بینی مدت‌زمان بستری شدن بیماران دستگاه گوارش
6	شناسایی و پیشگویی عارضه‌های بیماری	خوشه‌بندی	ترکیب الگوریتم ژنتیکی و K-Means	شناسایی و پیشگویی حملات قلبی
7	طب سنتی چینی	خوشه‌بندی	SVM درخت تصمیم Bayesian	کشف سندروم‌های مختلف
8	طب سنتی چینی	قوانین وابستگی	تحلیل شبکه پیچیده CNA	یافتن نقاط مطلوب طب سوزنی و الگوهای ترکیب گیاهان دارویی
9	طب سنتی چینی	خوشه‌بندی	ماشین بردار پشتیبان	پیشگویی شروع نفروپائی دیابت
10	طب سنتی چینی	خوشه‌بندی	شبکه‌های عصبی	درمان آرتیت روماتوئید

5- بررسی نمونه‌هایی از کاربردهای داده‌کاوی در پزشکی

5-1- تشخیص زودهنگام و یا پیشگیری از بیماری‌ها

Cheng و همکارانش به استفاده از الگوریتم‌های دسته‌بندی برای کمک به تشخیص زودهنگام بیماری‌های قلبی که نگرانی عمده‌ی بهداشت عمومی در سراسر جهان می‌باشد، اشاره‌کرده‌اند [7].

Cao و همکارانش استفاده از داده‌کاوی را به‌عنوان ابزاری برای کمک درروند نظارت در آزمایش‌های بالینی واکسن سرطان توصیف نموده‌اند. با کمک داده‌کاوی و تجسم ذهنی کارشناسان پزشکی توانستند الگوها و نابهنجاری‌های بهتری را در داده‌های جداول بیابند[8].

تشخیص بیماری دیابت نوع یک از دیگر تحقیقاتی است که با کمک تکنیک‌های داده‌کاوی انجام‌شده توسط آقای تهامی و همکارانش می‌باشد[28].

5-2- پیش‌بینی با کمک آنالیز و کشف دانش از داده‌های موجود

در یک تحقیق که Harris در دانشگاه بوستون در سال 2009 با استفاده از تکنیک‌های داده‌کاوی روی عوامل مؤثر در وزن کم نوزادان در هنگام تولد انجام داد و مدلی را برای پیش‌بینی این‌که وزن نوزاد یک مادر هنگام تولد کم خواهد بود را ایجاد نمود[9].

5-3- تشخیص غیرتهاجمی و پشتیبان تصمیم-گیری

برخی از روش‌های تشخیصی و آزمایشگاهی برای بیماران، پرهزینه، تهاجمی و دردناک هستند. Gorunescu در سال 2009 توضیح داد که چگونه تشخیص به کمک کامپیوتر و سونوگرافی به‌وسیله‌ی داده‌کاوی در تشخیص غیرتهاجمی سرطان به‌عنوان شیوه‌ی جدیدی می‌تواند به کار رود[10].

5-4- یافتن الگوهای پنهان در عوامل مختلف بیماری

در تحقیقی آقای کاج آبادی و همکارانش از داده‌کاوی برای به دست آوردن روابط مفید بین عوامل خطرزا در بیماری‌های قلب و عروق استفاده‌شده است. با اعمال داده‌کاوی روی‌داده‌های مربوط به1800 نفر در اصفهان مشخص شد مهم‌ترین متغیرهای مؤثر روی این عوامل خطرزا مربوط به LDL و استعمال دخانیات، تری گلیسیرید، سطح کلسترول، سن، شاخص توده‌ی بدن است[29].

5-5- پیش‌بینی میزان موفقیت روش‌های درمانی

لقاح آزمایشگاهی (IVF) یک روش کمکی برای پزشکان در تولیدمثل است که زوج نابارور را به یک حاملگی موفق می‌رساند. با توجه به غیرقابل‌پیش‌بینی بودن این کار، مقاله‌ای توسط Jurisica و همکارانش ارائه گردید که در آن با کمک داده‌کاوی و استخراج دانش از تجارب گذشته و کشف روابط بین داده‌ها یک سیستم مشاور پزشک در طی انجام کار بالینی طراحی گردید[11].

5-6- مضرات جانبی داروها

بعضی از داروها در ابتدا به‌عنوان داروهای غیر مضر برای انسان به تأیید رسیده‌اند، اما بعد از استفاده‌ی عموم در طولانی‌مدت اثرات زیانباری از خود را نشان می‌دهند. اداره‌ی دارو و غذای آمریکا با استفاده از داده‌کاوی به کشف دانش درباره‌ی عوارض جانبی داروها در پایگاه داده‌ی خود پرداخته است که با موفقیت 67 درصد عوارض جانبی داروها را 5 سال زودتر از شیوه‌ی سنتی شناسایی نموده است[12].

5-7- پیش‌بینی نتیجه‌ی تشخیص

Jonsdottir و همکارانش به ساخت مدلی برای پیش‌بینی نتیجه‌ی بروز سرطان سینه با کمک تکنیک‌های داده‌کاوی پرداختند. آن‌ها دریافتند که مدل مورداستفاده در پیش‌بینی از ارزیابی ریسک پزشکان از نتایج واقعی موفق‌تر عمل می‌کند[13].

5-8- کاربرد داده‌کاوی در زمینه بیماری قلبی

یکی از بزرگ‌ترین دلایل مرگ‌ومیر در جهان بیماری قلبی می باشد که سالانه تعداد زیادی از افراد را به کام مرگ می‌کشاند. لذا تشخیص به‌موقع این بیماری می‌تواند کمک بزرگی در حفظ سلامت جامعه و جلوگیری از مرگ‌ومیر و یا عوارض ناشی از آن باشد. در همین راستا، سودهاکار و همکاراش برای پیش‌بینی بیماری‌های قلبی، از تکنیک‌های داده‌کاوی استفاده نموده‌اند. الگوریتم‌های مورداستفاده آن‌ها، الگوریتم بیز، شبکه‌های عصبی، درخت تصمیم است [20].

5-9- کاربرد داده‌کاوی در زمینه سرطان

با توجه به بررسی انجام‌شده، 22.9 درصد از زنان در سراسر جهان به سرطان پستان رنج می‌برند. درواقع دومین عامل مرگ‌ومیر در بین زنان، سرطان و شایع‌ترین آن سرطان پستان است. پیش‌بینی به‌موقع این بیماری بسیار حائز اهمیت است زیرا می‌توان با درمان به‌موقع از بروز حوادث ناگوار مانند از دست دادن یک و یا هر دو سینه و یا فوت شخص درگیر اجتناب نمود. سومادیپ قوش و همکارانش با کمک دوطبقه بندی معروف، یعنی، با استفاده از شبکه عصبی پرسپترون چندلایه و ماشین بردار پشتیبان به‌پیش بینی این بیماری پرداخته‌اند که SVM دقتی برابر با 96.71٪ دارد، درحالی‌که MLP برابر با 95.71٪ می‌باشد [21].

5-10- کاربرد داده‌کاوری در زمینه دیابت

دیابت یکی از جدی‌ترین چالش‌های بهداشتی در کشورهای درحال‌توسعه و توسعه‌یافته است. با توجه به اعلام فدراسیون بین‌المللی دیابت، 285 میلیون نفر دیابتی در سراسر جهان وجود دارد. انتظار می‌رود که در 20 سال آینده این رقم به 380 میلیون نفر افزایش پیدا کند. با توجه به اهمیت موضوع، طراحی روشی برای تشخیص بیماری دیابت، با هزینه بهینه و عملکرد بهتر، نیاز این دوره می‌باشد. پایگاه داده دیابت هند، (پیما)، در آزمایشگاه یادگیری ماشین UCI یک استاندارد، برای داده‌کاوی در زمینه بیماری دیابت است. کوماری و چیترا نیز در مقاله خود از این پایگاه داده استفاده نموده‌اند .آن‌ها با استفاده از ماشین بردار پشتیبان (SVM)، که یک روش یادگیری ماشین برای طبقه‌بندی می‌باشد، به‌منظور تشخیص دیابت استفاده نموده‌اند .

دیتاست موردنظر شامل اطلاعات 768 بیمار زن می‌باشد و حداقل سن بیماران 21 سال می‌باشد. تمامی نمونه‌ها دارای ویژگی کلاس صفر و یا یک می‌باشند . صفر به معنی نداشتن بیماری دیابت و یک به معنی وجود بیماری می‌باشد. از کل دیتاست 268نفر در کلاس" 1 "و 500 مورد در کلاس" 0 "می-باشند. ویژگی‌های دیتاست موردنظر شامل : تعداد دفعات بارداری، غلظت گلوکز پلاسما در تست تحمل گلوکز خوراکی، مقدار انسولین تزریقی، شاخص توده بدنی، سابقه دیابت در خانواده، سن و...می‌باشد. بعد از پیاده‌سازی الگوریتم ذکرشده، برای ارزیابی مدل حاصله از ماتریس سردرگمی2 استفاده نمودند که بر اساس آن، مقدار صحت۱ و حساسیت۲ و ویژگی۳ مدل ارائه‌شده به ترتیب برابر است با 78٪، 80٪ و76.5٪ می‌باشد [22].

5-11- کاربرد داده‌کاوی در زمینه بیماران کلیوی

مراقبت و برقرار نگه‌داشتن دستیابی عروقی کارا برای بیماران همودیالیزی که مرحله نهایی بیماری کلیه هستند، از مسائل مهمی است که برای این دسته از بیماران کلیوی مطرح است . دو نوع دستیابی دائمی وجود دارد عبارت‌اند از: گرافت۴ و فیستول شریانی– وریدی (AVF)۵ که از ساختار عروق خود فرد بیمار است. احتمال شکست درروش دوم بسیار بالا است و عوامل مختلفی در این شکست دخیل‌اند. بامطالعه‌ای که با استفاده از تکنیک‌های با ناظر داده‌کاوی بر روی 137 مرد و 56 زن از بیماران مرکز کلیه هاشمی نژاد (بزرگ‌ترین بیمارستان کلیوی ایران در تهران) انجام شد ، از روش k-means و درخت تصمیم wj48و همین‌طور نرم‌افزار تحلیل وکا۶ استفاده شد، که نتایج حاصله نشان داده است که با درصدهای متفاوت چهار عامل مؤثر (دیابت، فشارخون، سیگار کشیدن و سن) از عوامل شکستAVF در بیماران همودیالیزی هستند، اما مهم‌ترین عامل سن می‌باشد[23].

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد که به تولید قاعده می‌پردازد. در ساختار درخت تصمیم، پیش‌بینی به‌دست‌آمده از درخت در قالب یکسری قواعد توضیح داده می‌شود و به‌صورت سلسله مراتبی می‌باشد. دسته‌بندی‌هایی که توسط درخت تصمیم ایجاد می‌شوند، از روی شباهت داده‌های ذخیره‌شده در پارامترهای پیش‌بینی کننده ، قابل انجام می‌باشد.

درک مدل ایجادشده توسط درخت تصمیم آسان است به همین دلیل گریکو و همکارانش برای پیش‌بینی شکست پیوند کلیه و ارزیابی اثر متقابل بین شاخص توده بدن۷(BMI) و دیگر عوامل خطرزا، از الگوریتم C 4.8 که گونه‌ای از درخت تصمیم می‌باشد، استفاده نموده‌اند. آن‌ها برای تحقیق خود 194 بیمار کلیوی، شامل 128 مرد و 66 زن که تحت عمل پیوند کلیه قرارگرفته بودند را به مدت 5 سال تحت بررسی قراردادند. متوسط سن بیماران موردبررسی، در زمان پیوند، سنی مابین 12/5 تا43/9 سال داشتند.

ویژگی‌های دیتاست جمع‌آوری‌شده شامل: سن، جنس، مدت‌زمان دیالیز، زمان پیوند، جنسیت اهداکننده، نوع اهداکننده(انسان زنده، مرگ مغزی)، سن اهداکننده،شاخص توده بدنی در زمان پیوند، عدم تطابق آنتی‌ژن‌های طرفین۸ (HLA)، تأخیر کارکرد گرافت۹ (DGF)، شدت پس‌زدگی۱۰ کلیه (ARE)پیوندی از سوی بدن شخص، ضایعات مزمن بعد از عمل۱۱ (CAN) است. و هر بیمار با یک کلاس مشخص‌شده است . در کل دیتاست شامل دو کلاس شکست۱۲ در برابر عدم شکست۱۳ می‌باشد . درنهایت آن‌ها توانستند با استفاده از الگوریتم طبقه‌بندی به تولید درخت تصمیم‌گیری برسند که بر اساس آن بتوانند به تأثیر DGF، CAN و BMI در نتایج پیوند کلیه برسند. بر اساس نتایج حاصله تمامی افراد پیوندی که در معرض خطر ARE, DGF و CANمی‌باشند، و همین‌طور دارای شاخص بدنی بیشتر از kg/m2 24.8 می‌باشند، درنهایت پیوند آن‌ها به شکست منتج خواهد شد . نتیجه اعتبار سنجی بر روی درخت حاصله نشان داد حساسیت 88.2٪ و ویژگی آن 73.8٪ می‌باشد[24].

تشخیص و درمان سنگ کلیه یکی از کارهای چالش‌برانگیز برای محققان در دهه حاضراست. این بیماری به علت تغییرات آب و هوایی رخ می‌دهد. در دهه حاضر، که هدف "پیشگیری بهتر از

[1] Accuracy
[2] Sensitivity
[3] Specificity
[4] Graft
[5] Arteriovenous Fistula
[6] Weka

[7] Body Mass Index
[8] Human Leukocyte Antigen
[9] Delayed Graft Function
[10] Acute Rejection Episode
[11] Chronic Allograft Nephropathy
[12] FAIL
[13] NO FAIL

www.kahrobaonline.ir

درمان است" می‌باشد، یافتن روشی مناسب برای تشخیص به‌موقع این درد بسیار حائز اهمیت می‌باشد.ازاین‌رو کالادهار و همکارانش با تحلیلی که با استفاده از دو الگوریتم درخت تصمیم C4.5 و ماشین بردار پشتیبان انجام دادند به نتایج خوبی در تجزیه‌وتحلیل و پیش‌بینی سنگ کلیه رسیدند. ویژگی‌های مانند وضعیت تأهل، مصرف چای و شیر، گروه خونی B، Rhمثبت، مسافرت، نوشیدن آب و... را برای وقوع بیشتر سنگ کلیه مستعد می‌کند دقت حاصل از دو الگوریتم مورداستفاده برابر با 93٪ برای الگوریتم درخت تصمیم و 91.98٪ برای ماشین بردار می‌باشد [25].

6- نتیجه‌گیری

امروزه در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف از اهمیت فراوانی برخوردار است. مراکز پزشکی با مقاصد گوناگونی به جمع‌آوری این داده‌ها می‌پردازند. تحقیق روی این داده‌ها و به دست آوردن نتایج و الگوهای مفید در رابطه با بیماری‌ها، یکی از اهداف استفاده از این داده‌ها است. داده‌کاوی روی داده‌های پزشکی از اهمیت بالایی برخوردار است. طراحی سیستم‌های کمک تصمیم جهت یاری‌رساندن به پزشکان در زمینه تشخیص نوع بیماری و یا انتخاب نوع درمان مناسب، پیش‌بینی میزان موفقیت روش‌های درمانی و مواردی از این قبیل با کمک داده‌کاوی می‌تواند کمک شایانی در زمینه نجات جان انسان‌ها انجام دهد.

امروزه باوجود عواملی چون حساسیت علم پزشکی و گره خوردن آن باجان انسان‌ها و حریم شخصی و محرمانگی داده‌های مرتبط با بیماران، داده‌کاوی به‌طور فزاینده‌ای برای رفع مشکلات متعدد در کشف دانش و در بخش پزشکی به کار گرفته‌شده است.

مراجع

[1] U. Fayyad, S G. Piatetsky, P. Smyth, R. Uthurusamy, "Advances in Knowledge Discovery and Data Mining", Massachusetts: MIT Press, 625 p, 1996.

[2] Wager K.A., Lee F.W., Glaser J.P., "Managing Health Care Information Systems: A Practical Approach for HealthCare Executives", New Jersey: John Wiley & Sons, 2005.

[3] A. Adel, P. Ahmadi, M. Sabet, "Desining Model for Choosing human resources with data mining approach. Journal of Iranian Technology, Vol. 2, No. 4, 2010 [In Persian].

[4] J. Han and M. Kamber, "Data Mining Concepts and Techniques", 2th edn, San Francisco: Morgan Kaufmann, 743 p, 2006.

[5] N. Wickramasinghe, J N. Gupta, S K. Sharma., "Creating Knowledge-Based Healthcare Organizations", Hershey: Idea Group Inc (IGI), 2005.

[6] K. LaTour and S. Eichenwald, "Health Information Management: Concepts, Principles", Chicago: AHIMA, pp. 478-480, 2002.

[7] T H. Cheng, C P.Wei, V S. Tseng , "Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches", Utah, Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06), 2006.

[8] X. Cao, K B. Maloney, V. Brusic, "Data Mining of Cancer Vaccine Trials: A Bird's-Eye View", Immunome Research, DOI:10.1186/1745-7580-4-7, 2008.

[9] Harris M., 2009, "Predicting Low Birth Weight Through Data Mining", Boston University, http://cs-people.bu.edu/dgs/courses/cs105/hall_of_fame/mpharris.pdf, [Accessed 22 May 2012]

[10] Gorunescu F., 2008, "Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection, International Journal of Biological and Life Sciences", vol. 4, No. 2, 2008.

[11] Jurisica I., Mylopoulos J., Glasgow J., Shapiro H., Casper R., 1998, "Case-Based Reasoning in IVF: Prediction and Knowledge Mining", Journal Artificial Intelligence in Medicine, Published Elesvier, vol. 12, p. 1–24

[12] Wilson A., Thabane L., Holbrook A , 2003, "Application of data mining techniques in pharmacovigilance", British Journal of Clinical Pharmacology, vol 57, issue 2, pp.127-134, 2003.

[13] Jonsdottir T., Thora Hvannberg T., Sigurdsson H., Sigurdsson S., 2008, The Feasibility of Constructing a Predictive Outcome Model for Breast Cancer Using the Tools of Data Mining, Journal Expert Systems with Applications, Published Elsevier, vol. 34 (2008), p. 108–118.

[14] Mehrpoor G,Azimzadeh M, Monfared A. "Data Mining: A Novel Outlook to Explore Knowledge in Health and Medical Sciences." International Journal of Travel Medicine and Global Health.2014;2(3);87-90.

[15] Oliyaei A, Salmasi N. "An efficient model to

predict duration of hospitalization of digestive system patients." 5th Conference of Data mining of Iran, 2011; Amirkabir University .
[16] Mohammadi M." Hybrid Algorithm for predict heart disease." International Research Journal of Applied and Basic Sciences, 2015,IRJOBS, ISSN 2251-838X / Vol, 9 (3).
[17] Pardalos PM, Tomaino V, Xanthopoulos P. "Optimization and data mining in medicine." 2009; Springer.
[18] Lin WT, Wu YC, Zheng JS, Chen MY. "Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital". Expert Systems with Applications, 2011; 38: 11078–11084.
[19] Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Guo Y, Zhang H, Gao Z, Yan X. "Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support." Artificial Intelligence in Medicine Journal, 2010; 48: 139–152.
[20] Sudhakar, K, and Dr M Manimekalai. "Study of Heart Disease Prediction using Data Mining." International Journal of Advanced Research in Computer Science and Software Engineering 4, no. 1 (2014): 1157-1160.
[21] Ghosh, Soumadip, Sujoy Mondal, and Bhaskar Ghosh. "A Comparative Study of Breast Cancer Detection based on SVM and MLP BPN Classifier." IEEE, 2014.
[22] Anuja Kumari, V, and R Chitra. "Classification Of Diabetes Disease Using Support Vector Machine." International Journal of Engineering Research and Applications 3, no. 2 (2013): 1797-1801.
[23] Brosius, FC. perspectives on systems bioloqy application in diabetic kidney disease. 2012.
[24] Greco, R, T Papalia, D Lofaro, S Maestripieri, and D Bonofiglio,R Mancuso. "Decisional Trees in Renal Transplant Follow-up." Transplantation Proceedings, 42. Elsevier, 2010. 1134–1136
[25] Kaladhar, DSVGK, Krishna Apparao Rayavarapu, and Varahalarao Vadlapudi. "Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis." 2012: 2-5.
[26] مشکانی ع، ناظمی ع، 1388، مقدمه ای بر داده کاوی، چاپ اول، مشهد، انتشارات دانشگاه فردوسی مشهد، 456 صفحه
[27] شرابیانی الف،1390، پایان‌نامه کارشناسی ارشد طراحی یک سیستم هوشمند برای پیش‌بینی بیماری قلبی با استفاده از تکنیک‌های داده‌کاوی، تهران، دانشگاه صنعتی شریف.
[28] تهامی الف، بامشکی م، خلیل زاده م، 1386، تشخیص بیماری دیابت نوع یک بااستفاده ازترکیب الگوریتم ANFIS و GA-NN، هشتمین کنفرانس سیستم های هوشمند ایران، مشهد، دانشگاه فردوسی مشهد، 9-7 شهریور 1386
[29] کاج آبادی ع، سرایی م، عسگری ص، 1388، راهکاری جهت کشف روابط بین عوامل خطرزای قلبی، سومین کنفرانس داده کاوی ایران، تهران، دانشگاه علم و صنعت ایران، 27-26 آبان 1388
[30] غضنفری م، علیزاده س، تیمورپور ب، 1390، داده کاوی و کشف دانش، چاپ دوم، تهران، انتشارات دانشگاه علم و صنعت ایران، 403 صفحه
[31] سپهری م، رهنما پ، شادپور پ، تیمورپور ب. 1388. ساخت مدلی برای تعیین نوع درمان سنگ حالب با راهکار داده-کاوی. مجله دانشکده پزشکی، دانشگاه علوم پزشکی تهران، صفحه 421-427