

## بررسی کاربردهای داده کاوی در نظام سلامت

لیلا غلامحسینی<sup>۱</sup>، مصطفی دامرودی<sup>۲\*</sup>

### چکیده

مقدمه: با گسترش داده‌های ذخیره‌شده در پایگاه‌های داده پزشکی نیاز به ابزارهای تخصصی برای دسترسی به داده‌ها، تجزیه و تحلیل داده‌ها، کشف دانش، و استفاده مؤثر از داده‌ها می‌باشد. داده کاوی یکی از مهم‌ترین این روش‌ها می‌باشد. در این مقاله تکنیک‌ها و نمونه‌هایی از کاربرد داده کاوی برای تشخیص پزشکی و پیش‌بینی بیماری‌ها در نظام سلامت بیان شده است.

مواد و روش‌ها: مطالعه حاضر با جستجوی مروری در متون موجود در پایگاه‌های اطلاعاتی فارسی و انگلیسی شامل **Magiran**، **Iranmedex**، **SID**، **Google Scholar**، **OVID**، **Scopus**، **PubMed** طی سال‌های ۱۹۹۸ تا ۲۰۱۳ انجام گردید.

یافته‌ها: داده کاوی به معنای جستجوی خودکار منابع داده‌ای بزرگ، جهت یافتن الگوها و وابستگی‌هایی است که تحلیل‌های ساده آماری قادر به انجام آن نیستند. یکی از زمینه‌هایی که نیازمند استفاده از این ابزارها جهت تحلیل داده‌های وسیع و مدل‌سازی پیش‌گویانه با روش‌های محاسباتی جدید است، علم پزشکی می‌باشد. در علم پزشکی کشف و تشخیص به موقع بیماری‌ها می‌تواند از ابتلا به بسیاری از بیماری‌های مهلک نظیر سرطان جلوگیری نموده و موجب نجات زندگی مردم گردد. این مطالعه نشان می‌دهد که پیشگویی‌های داده کاوی ابزارهای ضروری را برای محققان و پزشکان جهت بهبود در پیشگیری از بیماری‌ها، روش‌های تشخیصی و برنامه‌های درمانی فراهم می‌نمایند.

بحث و نتیجه‌گیری: امروزه در دانش پزشکی جمع‌آوری داده‌های بیماری‌های مختلف اهمیت فراوانی یافته است. پیشرفت‌های دهه اخیر در ارتباط با فناوری‌های اطلاعات و نرم‌افزار کمک شایانی به بررسی‌های همه‌جانبه و کامل‌تر از داده‌های حجیم تولید شده به عمل آورده و توانسته با استفاده از علوم گوناگون مثل آمار، کامپیوتر، یادگیری ماشینی و.... به جستجوی دانش نهفته در داده‌ها پرداخته و علم نوینی را به نام داده کاوی به وجود آورد.

کلمات کلیدی: داده کاوی (data mining)، سلامت (health)، کشف دانش (knowledge discovery)

### مقدمه

نمایان شده است. داده کاوی از جمله پیشرفت‌های فناوری در

راستای مدیریت داده است.

از سال ۱۹۵۰ به بعد که رایانه، در تحلیل و ذخیره‌سازی داده‌ها به کار رفت، حجم اطلاعات ذخیره‌شده در آن پس از حدود ۲۰ سال دو برابر شد و همزمان با پیشرفت فناوری اطلاعات، حجم داده‌ها در پایگاه داده‌ها هر دو سال یک‌بار، دو برابر شد و همچنان با سرعت بیشتری نسبت به گذشته حجم اطلاعات ذخیره‌شده بیشتر

امروزه در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف دارای اهمیت فراوانی است. مراکز پزشکی با مقاصد گوناگون به جمع‌آوری این داده‌ها می‌پردازند. تحقیق روی این داده‌ها و کشف نتایج و رابطه در مورد بیماری‌ها یکی از اهداف استفاده از این داده‌ها است. امروزه با توجه به ظهور نظام‌های اطلاعات یکپارچه و رشد فناوری اطلاعات این مهم بیش از پیش

۱- گروه مدیریت اطلاعات سلامت دانشگاه علوم پزشکی ایران، دانشگاه علوم پزشکی ارتش، تهران، ایران

۲- دانشکده پیراپزشکی، دانشگاه علوم پزشکی ارتش، تهران، ایران (\*نویسنده مسئول)

تلفن: ۰۹۱۵۴۷۶۳۸۷۸ آدرس الکترونیک: mostafadamroodi1992@yahoo.com

و بیشتر می شود.

شدت رقابت ها در عرصه های علمی از جمله سلامت به واسطه این هدف که (چگونه سازمان های سلامت هزینه ها را کاهش و کیفیت را افزایش دهند؟) اهمیت دسترسی به داده ها و اطلاعات را دو چندان کرده است، به عبارت دیگر در هر نوع برنامه بهبود کیفیت متمرکز بر بیمار داده ها قلب آن برنامه به حساب می آیند. (۱)

کشف دانش (Knowledge discovery) که بیش از یک دهه وارد محیط مالی و پزشکی شده است با هدف شناسایی سوءاستفاده های مالی وارد عرصه سلامت شد، اما به تدریج در حوزه بالینی نیز مورد استفاده قرار گرفت. صنعت سلامت به طور مستمر در حال تولید میزان زیادی داده می باشد و افرادی که با این نوع داده ها مواجه هستند دریافته اند که بین جمع آوری تا تفسیر داده شکاف وسیعی وجود دارد و داده کاوی از جمله شیوه هایی است که می تواند این صنعت را از تحلیل عمیق این داده ها بهره مند سازد و به توسعه تحقیقات پزشکی و تصمیم گیری های علمی در زمینه تشخیص و درمان منتج شود. (۲)

## مواد و روش ها

مطالعه حاضر با جستجوی مروری در متون علمی مرتبط موجود در پایگاه های اطلاعاتی انگلیسی و فارسی شامل، Magiran, Iranmedex, SID, Google Scholar, ovid, scopus, PubMed و minig knowledge health, discovery مربوط به سال های ۱۹۹۸ تا ۲۰۱۳ انجام شد.

## تعریف داده کاوی (Data Mining)

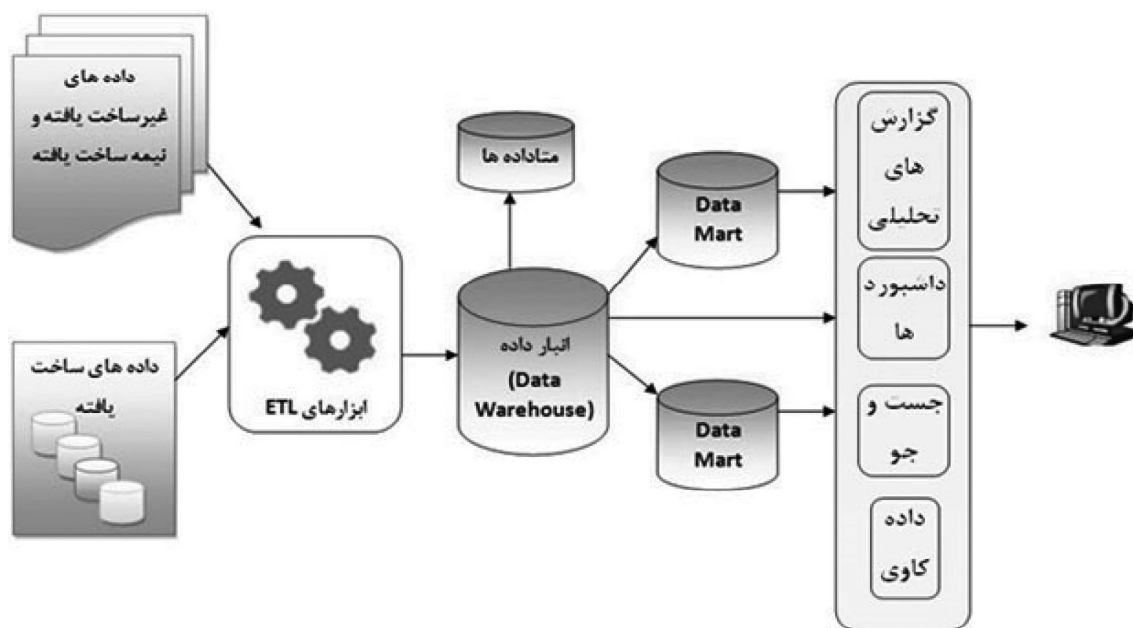
در دهه ی اخیر با تسهیل جمع آوری داده های پزشکی و دسترسی پذیری حجم زیاد داده های الکترونیکی، اکتشاف دانش توسعه یافته است. امروزه پژوهشگران از طریق فرایند اکتشاف دانش، اقدام به شناسایی روابط بین متغیرها، شناسایی الگوها و تولید مدل های پیش بینی پیامد بیماری ها می کنند. الگوریتم های یادگیری ماشینی از ابزارهای مرسوم در فرایند اکتشاف دانش بوده که با استفاده از آن ها مدل های پیش بینی با دقت بالایی ایجاد می شوند. به استفاده از الگوریتم های یادگیری ماشینی برای اکتشاف دانش در حجم عظیمی از داده ها، داده کاوی گفته می شود. (۳)

نگاهی به ترجمه لغوی داده کاوی به ما در درک بهتر این واژه کمک می کند. واژه لاتین Mine به معنای استخراج از منابع نهفته و با ارزش اطلاق می شود. ادغام این کلمه با Data به معنی داده، بر جستجوی عمیق از داده های قابل دسترس با حجم زیاد برای یافتن اطلاعات مفید که قبلاً نهفته بودند تأکید دارد. به عبارت دیگر داده کاوی استخراج اطلاعات مفید، مفهومی و ناشناخته از پایگاه داده می باشد. به لحاظ فنی، داده کاوی عبارت از فرآیندی است که در میان حوزه های گوناگون بانک های اطلاعاتی ارتباطی بزرگ، همبستگی ها یا الگوهایی را پیدا می کند. داده کاوی در حقیقت یکی از ابزارهای علم آمار به حساب می آید، اما گاهی دیده می شود که در برخی موارد آن را به عنوان شاخه ای جداگانه در نظر می گیرند و یا حتی سعی در بیان تفاوت های آن با آمار دارند. (۴)

روش های استخراج داده های سلامت عبارت اند از: تصویرسازی اطلاعات، تجزیه و تحلیل اطلاعات مرتبط، تجزیه و تحلیل منابع داده ها و طبقه بندی شبکه ها که تمام آن ها در پایگاه اطلاعاتی بیماران به کار گرفته می شود. این روش ها می توانند به شناسایی بیماران کمک کرده و فاکتورهای مهم موجود در بیماری را مشخص نمایند. در این ساختار فنون ساده تصویرسازی بیشترین استفاده را دارد که با دانش روز پزشکی سازگار است. (۵)

بیشتر افرادی که علم داده کاوی را مورد استفاده قرار می دهند افراد متخصصی در یک زمینه خاص علمی (پزشک، رادیولوژیست، مدیر فروش، ...) بوده که نه تنها به داده های خاص خود دسترسی دارند بلکه به جمع آوری آن نیز اقدام می ورزند. این افراد ترجیح می دهند داده های خود را بهتر بشناسند و همچنین مایل اند دانش جدیدی را در رابطه با زمینه فعالیت خود کشف نمایند.

در علم پزشکی کشف و تشخیص به موقع بیماری ها می تواند از ابتلا به بسیاری از بیماری های مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد. امروزه با پیشرفت های بیولوژیکی، توسعه تکنولوژی و استفاده از فناوری های روز و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع آوری اطلاعات دقیق تری در مورد بیماران بپردازند که تحلیل آن ها به دلیل حجم بالای اطلاعات مشکل بوده و نیازمند فناوری های نوین می باشد که تکنولوژی های داده کاوی به کمک الگوریتم های قدرتمند خود به این مهم دست یافته است. هدف از روش های پیشگویی داده کاوی در پزشکی



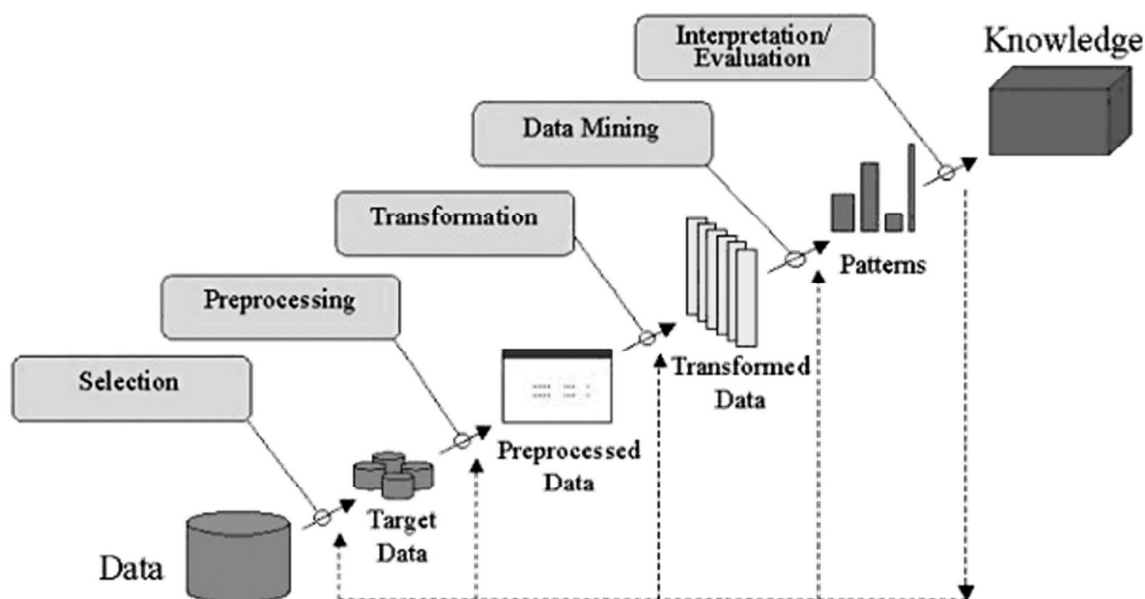
داده کاوی در واقع مرحله‌ای از فرآیند کشف دانش تلقی می‌شود. مراحل موجود در فرآیند استخراج دانش در شکل یک نشان داده شده است و شامل مراحل زیر است:

- **جمع‌آوری داده‌ها:** در این مرحله پس از پالایش داده‌ها، چندین منبع داده‌ای در یک انبار داده یکپارچه قرار می‌گیرند.
- **انتخاب و آماده‌سازی داده‌ها:** در این قسمت داده‌های مرتبط انتخاب می‌گردد و به شکل مناسبی برای داده کاوی تبدیل می‌شوند.

بالینی ساخت یک مدل پیش‌گویانه است که به پزشکان کمک می‌کند تا روش‌های پیشگیری، تشخیص و برنامه‌های درمانی خود را بهبود بخشند. (۶)

#### مراحل داده کاوی

به صورت ساده می‌توان بیان نمود که داده کاوی به استخراج دانش از حجم انبوهی از داده‌ها اطلاق می‌شود. بسیاری از افراد این واژه را مترادفی برای واژه‌ی کشف دانش می‌دانند، اما با توجه به شکل زیر



شکل ۱- مراحل داده کاوی و کشف دانش از داده

بیمه‌ای یک موسسه سلامت، نمونه‌ای از استراتژی یادگیری نظارت شده است؛ در این استراتژی مدل‌ها و ویژگی‌ها برای ما شناخته شده و با هدف پیش‌بینی داده‌ها و کشف اطلاعات به کار می‌رود، اما در شیوه یادگیری فاقد نظارت، ویژگی‌ها و مدل‌های خطاهای مطالبات شناخته شده نیست، اما الگوها و خوشه‌های حاصل از داده کاوی منجر به کشفیات جدید می‌شود. (۹)

### رایج‌ترین تکنیک‌های داده کاوی و کاربرد آن در حوزه سلامت

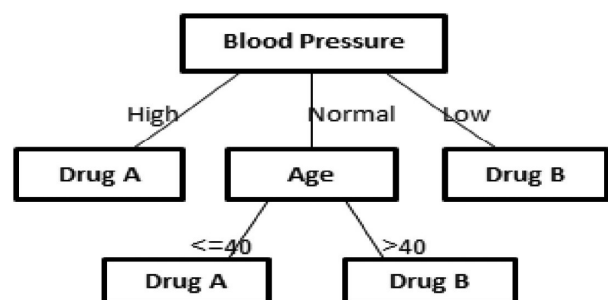
#### تکنیک‌های داده کاوی

داده کاوی از تکنیک‌هایی همچون مدل‌های Bayesian، درخت تصمیم‌گیری، شبکه‌های عصبی مصنوعی، قواعد ارتباطات و الگوریتم‌های ژنتیک برای کشف الگوها استفاده می‌کند که برخی از آن‌ها بیان می‌شود.

– **درخت تصمیم‌گیری (Decisions Tree):** این تکنیک ساختاری شبیه به درخت دارد و مجموعه قواعدی که منجر به تصمیم‌گیری شده را توصیف می‌کند و سهولت تفسیر از ویژگی‌های مهم آن است. برای مثال درخت تصمیم‌گیری می‌تواند عوامل مؤثر بر میزان بقای پیوند کلیه را تعیین کند و همچنین استفاده از درخت الگوریتم DRG (Diagnosis Related Groups)

(کدهای تشخیصی وابسته) در بازپرداخت هزینه‌های بیمه پزشکی سالمندان در ایالات متحده مثال کلاسیکی از این تکنیک است. شکل ۲-نمونه‌ای از درخت تصمیم‌گیری را برای مثالی از داده‌های پزشکی نشان می‌دهد.

در درخت بالا می‌بینیم که چگونه یک پزشک می‌تواند بر اساس صفات خاصی فشار خون و سن بیمار داروی مناسب جهت مداوای



شکل ۲- درخت تصمیم‌گیری

- **داده کاوی:** فرآیندی که با به خدمت گرفتن روش‌های هوشمند در میان داده‌ها به دنبال الگوهای خاصی می‌گردد.
- **تفسیر و ارزشیابی الگوها:** از میان انبوهی از الگوها، با تعریف معیارهای متنوع الگوهای محدودی برای تفسیر و تحلیل انتخاب می‌شوند.
- **ارائه دانش:** در این مرحله با کمک ابزار بصری سازی و فن‌های مختلف داده کاوی، دانش کشف‌شده به کاربر و یا تحلیل‌گر ارائه می‌شود. (۷)

### مهم‌ترین دلایل استفاده از داده کاوی (Wickramasinghe) و همکاران

- ۱- دلیل اصلی استفاده از داده کاوی، افزایش حجم داده‌های موجود و آتی سازمان‌ها است که نیاز به پردازش فراتر از رویکردهای سنتی دارد. برای انسان کنار آمدن با این حجم انبوه و در حال رشد داده‌ها با اتکای محض به شیوه‌های سنتی غیرممکن است.
- ۲- ذهن انسان به هنگام تحلیل انبوه داده‌ها ممکن است دچار خطا شود و علاوه بر این همیشه سعی بر این دارد که نتایج را مبنی بر تجربیات و آزمون قبلی بنا نهد، اما داده کاوی بدون تکیه بر فرضیه‌های پیشین به تحلیل داده‌ها می‌پردازد.
- ۳- یکی از مهم‌ترین مزایای داده کاوی به ویژه در حجم وسیع داده‌ها این است که داده کاوی نسبت به استخدام تیمی از افراد خبره بسیار کم هزینه‌تر است.

داده کاوی از ساخت مدل‌های تحلیلی، دسته‌بندی و پیش‌بینی اطلاعات و ارائه نتایج با استفاده از ابزارهای مرتبط استفاده می‌کند. برای اینکه الگوریتم داده کاوی بتواند عمل استخراج دانش را به خوبی انجام دهد، نیاز به یک سری پیش‌پردازش‌ها بر روی داده‌ها و یک سری پس‌پردازش‌ها بر روی الگوهای استخراج شده دارد. (۸)

### استراتژی‌های داده کاوی

به طور کلی هدف داده کاوی، یادگیری و آموختن از داده‌ها است و بر این اساس دو دسته کلی از استراتژی‌های داده کاوی شامل یادگیری نظارت شده و یادگیری فاقد نظارت وجود دارد. شیوه‌های یادگیری نظارت شده زمانی به کار می‌رود که ارزش متغیرهای ورودی برای ما شناخته شده باشد. یافتن مدل‌های پیش‌بینی خطا در مطالبات

موارد است. گورانسکو (Gorunescu) به نقل از کانلاس (Canlas) نشان داد که داده کاوی می تواند به ادغام CAD (Computer Assisted Diagnosis) (تشخیص به کمک رایانه) و اندوسکوپی اولتراسونوگرافی در تشخیص غیر تهاجمی سرطان به عنوان یک شیوه جدید به کار رود.

**۳- داده کاوی در کنترل عفونت بیمارستانی:** در ایالات متحده سالانه دو میلیون نفر به عفونت های بیمارستانی مبتلا می شوند، لذا تمرکز زیادی برای شناسایی این بیماران صورت گرفته است. به عنوان مثال در ایالت آلباما نوعی سیستم نظارتی وجود دارد که از تکنیک های داده کاوی استفاده می کند این سیستم با استفاده از قوانین و روابط داده کاوی بر روی کشت خون بیمار و داده های بالینی به دست آمده از سیستم اطلاعات آزمایشگاه (Laboratory Information System (LIS) الگوهای جدید و جالب توجهی را مشخص می سازد و ماهانه الگوهایی که توسط کارشناسان کنترل عفونت مورد بررسی قرار می گیرد را تهیه می کند. سازندگان این سیستم دریافته اند که ارتقای کنترل عفونت با سیستم داده کاوی حساس تر از سیستم کنترل عفونت سنتی عمل می کند. (۱۱)

**۴- داده کاوی در تعیین نوع درمان:** به کارگیری داده کاوی بر روی داده های پزشکی دستاوردهای حیاتی و اثرگذاری را در انتخاب نوع درمان مناسب و نجات جان انسان ها به ارمغان آورده است، به عنوان مثال در بیمارستان شهید هاشمی نژاد تهران برای تعیین نوع درمان سنگ حالب از راهکار داده کاوی استفاده می شود. در این بیمارستان یک الگوریتم درختی وجود دارد که پزشک بر اساس آن درمانی که میزان موفقیت بالاتری برای بیمار دارد را انتخاب می کند.

**۵- داده کاوی در پرونده الکترونیک سلامت:** در حال حاضر مطالعات متعددی مؤکد این مطلب است که تکنیک های داده کاوی ابزار مؤثری را برای شناسایی الگوهای مهم سلامت از درون پرونده های پزشکی فراهم می کنند.

پرونده های سلامت رایانه ای به واسطه دربرداشتن مجموعه ای از داده ها درباره تشخیص، درمان، اقدامات آزمایشگاهی و دارویی به طور بالقوه منبع غنی از دانش هستند و کشف دانش از انبوه داده به کمک داده کاوی صورت می گیرد.

او را تجویز کند. این مثال به خوبی نشان می دهد که چگونه یک درخت تصمیم برای نمایش یک مدل طبقه بندی استفاده می شود. (۷)

**- شبکه های عصبی (Neural Networks):** این تکنیک مدل های پیش بینی غیر خطی تولید می کند که یاد می دهد چگونه یک الگو با یک پروفایل خاص قابل تطبیق است، اما درباره علت رسیدن به این نتیجه خاص توضیحی ارائه نمی کند. برای مثال شبکه های عصبی قادرند مشخص کنند که چه نوع بیماری هایی احتمال دارد با یک بیماری همراه شود و با تحلیل تصاویر، نوارهای قلبی و سایر مشاهدات بالینی به تشخیص، درمان و تولید دارو کمک نمایند.

**- الگوریتم های ژنتیک (Genetic Algorithms):** تکنیک های بهینه ای برای ارتقای سایر الگوریتم های داده کاوی، به گونه ای که از بهترین مدل بر روی مجموعه داده ها استفاده کنند و می تواند برای یک بیماری خاص بهترین برنامه درمانی را تعیین کند.

### نمونه هایی از کاربرد داده کاوی در پزشکی

**۱- کاربرد داده کاوی در بیماری های قلبی و عروقی:** رشد چشمگیر بیماری های قلبی و عروقی، اثرات و عوارض آن ها و هزینه های بالایی که بر جامعه وارد می کند، باعث شده که جامعه پزشکی به دنبال برنامه هایی جهت بررسی بیشتر، پیشگیری، شناسایی زود هنگام و درمان مؤثر آن باشند. از این رو با استفاده از داده کاوی و کشف دانش در سیستم مراکز قلب و عروق می توان دانش ارزشمندی را ایجاد کرد که این دانش کشف شده می تواند باعث بهبود کیفیت سرویس به وسیله مدیران مرکز شود و همچنین می تواند به وسیله پزشکان استفاده شود تا رفتار آینده بیماران قلبی و عروقی را از روی سابقه داده شده پیش بینی کند. تشخیص بیماری قلبی از روی ویژگی های گوناگون و نشانه ها، ارزیابی فاکتورهای ریسکی که باعث افزایش حمله قلبی می شود از مهم ترین کاربردهای داده کاوی و کشف دانش در سیستم بیماران قلبی و عروقی است. (۱۰)

**۲- داده کاوی در تشخیص های غیر تهاجمی:** برخی از اقدامات تشخیصی و آزمایشگاهی برای بیماران، تهاجمی و هزینه بر و در عین حال رنج آور هستند. به عنوان مثال بافت برداری از گردن رحم به منظور تشخیص سرطان گردن رحم از جمله این

ندارند را تحت پوشش بیمه تهیه می‌کنند. یکی از کاربردهای داده کاوی تشخیص این گونه سوء استفاده‌ها می‌باشد.

#### ۱۰- مطالعه انجام شده در مورد پیش‌بینی عوارض دیابت: با

توجه به اینکه بیماری دیابت، به عنوان یک بیماری بسیار مزمن شناخته شده است و آسیب‌های جبران ناپذیری به اندام‌ها و اعضاء حیاتی بدن وارد می‌کند، استفاده از ابزارهای هوشمند داده کاوی می‌تواند برای بهبود روش‌های شناسایی و کنترل بیماری به پزشکان کمک بزرگی باشد. در تحقیق انجام شده با استفاده از الگوریتم‌های داده کاوی به دسته‌بندی بیماران دیابتی بر اساس عارضه مشاهده شده در آن‌ها پرداخته شد. عوارض این بیماری را بر اساس دو دسته میکرواسکولار و ماکرواسکولار دسته‌بندی شد. از بین الگوریتم‌های داده کاوی بهترین نتایج از الگوریتم درخت تصمیم به دست آمد که دقت مدل آن برابر ۸۹،۰۶ درصد و صحت مدل ۸۹،۷۴ درصد است. بیشترین پارامترهای تأثیرگذار بر روی عوارض بیماری میزان فشار سیستولیک، سن، سابقه خانوادگی و چربی مضر شناخته شد. با استفاده از قوانین ایجاد شده بر روی یک نمونه جدید با ویژگی‌های مشخص، می‌توان پیش‌بینی کرد که این فرد احتمالاً دچار چه نوع عارضه‌ای خواهد شد. با کنترل عوامل تأثیرگذار بر بروز عارضه در تر بیمار، می‌توان امیدوار بود از بروز عارضه تا حدی اجتناب شود و یا آن را به تعویق انداخت. (۱۳)

#### ۱۱- مطالعه انجام شده در مورد پیش‌بینی عوارض عروق کرونر:

یکی از شایع‌ترین بیماری‌های قلبی عروقی، بیماری عروق کرونر است که بر اساس اطلاعات حاصل از یک مطالعه اپیدمیولوژیک که به بررسی علت مرگ و میر در سال ۱۳۸۸ پرداخته است، ۲۵،۶ درصد عامل منجر به مرگ محسوب می‌شود. بهترین روش ارزیابی بیماری عروق کرونر، آنژیوگرافی می‌باشد. با این وجود آنژیوگرافی یک روش گران و تهاجمی (Invasive) بوده و همراه با ریسک‌هایی همچون مرگ، سکته قلبی و مغزی می‌باشد. از این رو وجود نظام‌های پشتیبان تصمیم‌گیری در کنار روش‌های قبل از آنژیوگرافی برای کم کردن نتایج کاذب لازم به نظر می‌رسد. این سیستم‌ها با استفاده از تکنیک‌های داده کاوی می‌توانند به کشف الگوها در داده‌های پزشکی پرداخته و فرآیند تصمیم‌گیری را بهبود بخشند. در مطالعه انجام

۶- داده کاوی در رتبه‌بندی بیمارستان‌ها: رتبه‌بندی بیمارستان‌ها و برنامه‌های بهداشتی می‌تواند بر مبنای اطلاعات گزارش شده توسط ارائه دهندگان مراقبت باشد، بنابراین گزارش دهی استاندارد برای مقایسه معنی‌دار بیمارستان‌ها و رتبه‌بندی آن‌ها بسیار مهم است. از جمله شیوه‌های استاندارد کردن این گزارشات استفاده از داده کاوی است. به عنوان مثال اگر کدهای ICD با تکنیک‌های داده کاوی مانند خوشه‌بندی همراه شود، می‌تواند به ایجاد گزارشاتی که با نرخ واقعی میزان ناخوشی‌ها، مرگ و میر و سایر شاخص‌های کیفیت در رتبه‌بندی بیمارستان‌ها بیانجامد. (۸)

#### ۷- تعیین نوع رفتار با بیماران و پیشگویی میزان موفقیت اعمال

##### جراحی (۱۲)

۸- مدیریت درمان: با استفاده از داده کاوی می‌توان بیماری‌های مزمن و خطرناک را شناسایی کرد و نتیجه‌ی این کار مدیریت بهتر بر روی بیماران می‌باشد که می‌توانیم به این بیماری‌ها اولویت بیشتری اختصاص داده و در اورژانس‌ها و موقعیت‌هایی که زمان و منابع مورد نیاز در برخورد با تمام بیماران محدود می‌باشد، این بیماران را در اولویت اول قرار داد. حتی می‌توان تعداد موارد بستری در بیمارستان‌ها را کاهش داد و از پذیرش‌های بی‌جهت جلوگیری کرد. همچنین می‌توان عوارض دارویی و تداخلات دارویی در برخی از درمان‌ها را شناسایی کرد و برای درمان دیگر بیماران از نتیجه این الگوها استفاده نمود. مانند:

##### - ObamaCare: یک طرح برای اصلاح نظام خدمات درمانی و بیمه، که یکی از اصلاحات آن بهبود کیفیت درمانی بیمارستان‌ها، صرفه‌جویی در هزینه درمان و... است.

- Arkansas Data Network: به دنبال مراجعات و بستری مجدد، بهره‌برداری از منابع و مقایسه داده‌ها با انتشارات علمی پژوهشی در جهت تعیین بهترین گزینه درمان.

- Group Cooperative Health: انواع بیماران را طبق وضعیت پزشکی و دسترسی به منابع گروه‌بندی می‌کند و در نتیجه برای آموزش این گروه‌ها و جلوگیری یا مدیریت وضعیت بیماری بیماران، برنامه‌ریزی می‌کند.

۹- تشخیص قلب و سوءاستفاده: برخی از افراد سودجو در خرید دارو از بیمه سوء استفاده می‌کنند و داروهایی که خود نیاز



کاوی می‌باشد. داده‌های پزشکی نیز شامل بیماری‌ها و اطلاعات حساس افراد می‌باشند که این گونه اطلاعات در فرآیند داده کاوی نباید فاش شوند. مکانیزم‌ها و مدل‌هایی برای حفظ حریم خصوصی در داده کاوی طراحی شده که می‌توانند از فاش شدن اطلاعات حساس جلوگیری به عمل آورند. این مدل‌ها و روش‌ها شامل دستکاری داده‌ها، تصادفی سازی (Randomization) نمونه گیری و حفظ گمنامی (Maintain Anonymity) می‌باشد. اما اگر کارایی داده کاوی از اهمیت بیشتری برخوردار باشد، سیاست‌های امنیتی در جهت حفظ حریم خصوصی باید به همراه داده کاوی انجام شود که شامل دستکاری در الگوریتم‌های داده کاوی برای محدودسازی داده کاوی و همچنین جلوگیری از خرابکاری از طریق حسابرسی پرس و جو (query auditing) می‌باشد. (۶۱)

### جایگاه داده کاوی در ایران و سایر کشورهای جهان

در سال‌های اخیر علم داده کاوی محبوبیت خاصی در میهن عزیزمان ایران پیدا کرده است. حال می‌توان به برخی موارد از پروژه‌های اجرا شده و قابل اجرا در زمینه داده کاوی و دستاوردهای حاصل از آن در سازمان بیمه درمانی تأمین اجتماعی اشاره نمود:

- استفاده از تکنیک خوشه‌بندی و دسته‌بندی برای تعیین ویژگی مؤسسات با بدهی بالا، شناخت عملکرد کارفرمایان؛ استراتژی و تصمیمات آگاهانه با درجه اطمینان بالا جهت وصول بدهی‌ها که توسط مدیران ارشد سازمان صورت می‌پذیرد.
- پیش‌بینی و تخمین در داده کاوی به منظور پیش‌بینی الگوی رشد بدهی کارفرمایان، میزان درآمد، وصولی و هزینه‌های مرتبط با مدیریت امور مالی و ذخایر سازمان و همچنین در برآورد بودجه‌های مالی بسیار اثربخش خواهد بود.
- ارائه خدمات کوتاه مدت و بلندمدت سازمان با در نظر داشتن اصول مدیریت ارتباط با مشتریان (CRM) (Customer Relationship Management) با استفاده از فن‌آوری داده کاوی جهت جلب رضایت بیشتر مخاطبین (بیمه‌شدگان، مستمری بگیران و کارفرمایان) با شناخت بیشتر از داده‌های موجود میسر خواهد بود.
- کشف جرم و تقلب در سطح نظام‌های اطلاعاتی در دو حوزه بیمه‌ای و درمان، با استفاده از کاربردهای مؤثر داده کاوی و

شده برای پیش‌بینی بیماری عروق کرونر از تکنیک قدرتمند شبکه‌های عصبی استفاده گردید که مدل نهایی به دست آمده دارای دقت ۷۴/۱۹ درصد، ویژگی ۳۳/۲۵ درصد و حساسیت ۹۲/۴۱ درصد بود و توانست علاوه بر توانایی بالا در تشخیص افراد بیمار، تعداد قابل قبولی از افرادی که فاقد بیماری عروق کرونر بودند را نیز شناسایی کند. (۱۴)

### ۱۲- مطالعه انجام شده در مورد کاربرد داده کاوی در پیش‌بینی مرگ بیماران سوختگی:

پیش‌بینی پیامد بیماری سوختگی بر اساس شرایط بیمار از اهمیت بالایی برخوردار است. از آنجایی که پیش‌بینی صحیح پیامد بیماری‌ها ابزار مناسبی برای متخصصین جهت تصمیم‌گیری بالینی آگاهانه به حساب می‌آید، پژوهشگران پزشکی همیشه به دنبال مدل‌هایی جهت پیش‌بینی پیامد بیماری‌ها و تعیین اهمیت فاکتورهای خطر بوده‌اند. در این مطالعه با استفاده از ابزارها و الگوریتم‌های داده کاوی از جمله مدل شبکه عصبی مصنوعی که دقت بالایی در پیش‌بینی پیامد بیماران سوختگی داشت، استفاده شد. در نهایت می‌توان گفت پیش‌بینی صحیح، تأثیر قابل توجهی در تخصیص بهینه منابع بیمارستانی داشته و در نتیجه باعث ارتقای کیفیت ارائه خدمات مراقبت می‌گردد. (۱۵)

### چالش‌های داده کاوی در پزشکی:

اکثر روش‌های داده کاوی استاندارد، می‌توانند تنها الگوها را نمایش دهند، و قادر نیستند الگوها و تمایلات را شرح دهند. در مقابل، پزشکی نیاز به توصیف الگوها دارد زیرا یک تفاوت اندک می‌تواند نتیجه (مرگ یا ادامه زندگی) را تغییر دهد. مانند تشخیص بیماری آنفلوآنزا و آنتراکس (Anthrax) که نشانه‌های هر دو بیماری شبیه به هم هستند.

آماده کردن داده‌ها یا همان مرحله پیش‌پردازش (Preprocessing) داده‌ها یک مرحله بسیار دشوار است، چون داده‌ها باید کامل و بی‌نقص باشند تا داده کاوی بر روی داده‌ها درست انجام شود. چالش مهمی که با آن روبرو هستیم، کامل نبودن داده‌های پزشکی، استفاده از منابع مختلف داده‌ای و همچنین تلفیق اطلاعات و آماده‌سازی داده‌ها که امری بسیار دشوار است.

**حفظ حریم خصوصی در پزشکی:** موضوع حفظ حریم خصوصی موضوعی بسیار مهم در امر انتشار و تحویل داده‌ها پیش از داده

برنامه‌های درمانی خود را بهبود بخشند. نکته آخر اینکه استفاده از مدل‌های پیش‌بینی، هیچ‌گاه جایگزینی برای تصمیم‌گیری‌های پزشکان نمی‌باشد، بلکه هدف، فراهم‌سازی اطلاعات تکمیلی و پشتیبان برای ایشان در اتخاذ تصمیمات آگاهانه‌تر است.

### نتیجه‌گیری

با توجه به اهمیت و حساسیت داده‌های پزشکی و همچنین نیاز به حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد، در این مطالعه کاربردهای داده‌های پزشکی در حوزه سلامت مورد بررسی قرار گرفت.

به دنبال تولید مستمر و انبوه داده در سازمان‌های مراقبت سلامت و توسعه پایگاه‌های داده نسبت به دهه‌های گذشته نیازهای جدیدی را مانند خلاصه‌سازی خودکار داده و استخراج اطلاعات ذخیره‌شده از داده‌های خام به وجود آورده است که داده‌های پزشکی نمونه‌ای از آن می‌باشد. داده‌ها در عصر امروزی عمده‌ترین دارایی سازمان‌های سلامت هستند و موفقیت آن‌ها منوط به جمع‌آوری، ذخیره و تحلیل این داده‌ها است. از این رو بسیاری از سازمان‌ها به داده‌های پزشکی روی آورده‌اند، چرا که به واسطه داده‌های امکان کشف روابط و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین میسر خواهد شد.

داده‌های پزشکی در پزشکی و بیولوژی بخش مهمی از انفورماتیک زیست-پزشکی است و یکی از کاربردی‌ترین علوم رایانه در این علم بوده که در بیمارستان‌ها، کلینیک‌ها، آزمایشگاه‌ها و مراکز تحقیقاتی به کار گرفته شده است. امروزه سنجش سلامت بیشترین نیاز را به داده‌های پزشکی پیدا کرده است و حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد از جمله مواردی است که می‌تواند مؤکد این امر باشد.

### پیشنهادهای

در محیط رقابتی امروز، سازمان‌هایی که به واسطه استفاده از فن‌آوری‌های نوین همچون داده‌های پزشکی بتوانند داده‌ها را در راستای بهبود کیفیت سلامت به کار ببرند سریع‌تر به قله‌ی موفقیت خواهند رسید، لذا پیشنهاد می‌گردد تا سازمان‌های سلامت از این عرصه بازمانند.

بسیاری از مراکز تحقیقاتی کشورمان دارای حجم زیادی از داده‌ها هستند که یا هرگز تحلیل نمی‌شوند و یا اگر هم تحلیل و به دانش

تکنیک‌های کشف ناهنجاری‌ها و بررسی موارد پرت (Noise) قابل انجام می‌باشد. برای نمونه، می‌توان به استخراج الگوی مصرف داروهای خاص بر اساس سوابق داده‌های قبلی پرداخت و موارد ناهمگون و پرت را شناسایی نمود. علم داده‌های پزشکی در کشورهای توسعه‌یافته پیشرفت بیشتری داشته و به دنبال آن کارهای مرتبط با داده‌های پزشکی نیز گسترش بسیار زیادی پیدا کرده‌اند.

سایت Kdnugget یکی از مشهورترین پایگاه‌های اینترنتی است که در خصوص داده‌های پزشکی فعالیت می‌نماید. طی نظرسنجی صورت گرفته از این سایت در سال ۲۰۱۲، مشهورترین و پرکاربردترین حوزه‌های داده‌های پزشکی مشخص شد که حوزه بهداشت با ۱۷ درصد در جایگاه سوم قرار داشت. (۱۷)

### بحث

امروزه جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف در علم پزشکی اهمیت بسیاری یافته و مراکز پزشکی با مقاصد گوناگون این داده‌ها را گردآوری می‌کنند. مطالعه بر روی این داده‌ها و کشف نتایج و روابط بیماری‌ها یکی از اهداف جمع‌آوری این داده‌ها است. در علم پزشکی، کشف و تشخیص به موقع بیماری‌ها می‌تواند از ابتلا به بسیاری از بیماری‌های مهلک نظیر سرطان جلوگیری نموده و باعث نجات زندگی مردم گردد.

با پیشرفت‌های بیولوژیکی، تکنولوژی و به‌کارگیری فناوری‌های نوین و تجهیزات مدرن پزشکی، متخصصین قادرند تا به جمع‌آوری داده‌های دقیق‌تر بیماران بپردازند. به دلیل حجم بالای داده‌ها تحلیل آن دشوار بوده و نیازمند فناوری‌های جدید می‌باشد، که تکنولوژی‌های داده‌های پزشکی با کمک الگوریتم‌های قدرتمند خود به این مهم دست یافته است.

کشف دانش از پایگاه‌های داده پزشکی به منظور تشخیص مؤثر در پزشکی بسیار مهم است. هدف از داده‌های استخراج دانش از اطلاعات ذخیره‌شده در پایگاه داده و ایجاد شرح روشن و قابل فهم از الگوها است. از بررسی و مقایسه مطالعات صورت گرفته، به صورت کلی می‌توان نتیجه گرفت که هدف از روش‌های پیشگویی داده‌های پزشکی بالینی ساخت یک مدل پیش‌گویانه است که به پزشکان کمک می‌کند تا روش‌های پیشگیری، تشخیص و



گام‌های جدیدی را در پیش‌گیری، تشخیص، درمان و ارائه خدمات با کیفیت به مشتریان سلامت بردارند.

منتج شوند، به واسطه استفاده از شیوه‌های سنتی، امری مقطعی و زمان‌بر هستند؛ حال آنکه با روی آوردن به داده کاوی و اجرای آن می‌توانند داده‌ها را به ابزاری نیرومند و رقابتی تبدیل نموده و

## References

- 1- Rogers G, Joyner E. Mining Your Data for Healthcare Quality Improvement [Online]. 2011 [cited 2011 Aug 8]; Available from: URL: <http://www2.sas.com/proceedings/sugi22/EMERGING/PAPER139.PDF/>
- 2- Englehardt SP, Nelson R. Health care informatics: an interdisciplinary approach. Philadelphia: Mosby; 2002. pp125.
- 3- Koh HC, Tan G. Data mining applications in healthcare. J Health In Manage 2005; 19 (2): 64-72.
- 4- Basic Epidemiology-file type ppt → Payman Salamati, MD, MPH, Assistant Professor of community Medicine Tehran University of Medical Sciences 1387.
- 5- Medical Informatics, The internet Medicine ISSN 1436-9238, Print /ISSN 1464-5238 online c 2000 Taylor & Francis Ltd. Available from: <http://www.tandf.co.uk/journals>.
- 6- Hassanzadeh M, Razavi Ebrahimi SL. Comparison Classification of Data Mining Algorithms in Medical Sciences.
- 7- Esmaeili M. Concepts & Techniques datamining, Tehran: Niasedanesh Publisher; 1391.
- ۸- مقدسی ح، حسینیان ف، اسدی ف، جهانبخش م. داده کاوی و کاربرد آن در سلامت. مجله مدیریت اطلاعات سلامت. ۱۳۹۱، ۹ (۲): ۲۹۷-۳۰۴.
- 9- Obenshain MK. Application of data mining techniques to healthcare data. Infect Control. HospEpidemiol. 2004;25 (8): 690-5.
- ۱۰- کاج آبادی س، سرایی م، عسگری ص. داده کاوی پزشکی: راهکاری جهت کشف روابط بین عوامل خطرزای قلبی.
- ۱۱- نوروزی ف، طائفی همراه ع. مروری بر داده کاوی و بررسی شبکه‌های عصبی در چند زمینه‌ی مختلف.
- 12- Wickramasinghe N, Gupta JN, Sharma SK. Creating Knowledge-Based Healthcare Organizations. Hershey: IdeaGroup Inc (IGI). 2005.
- ۱۳- عامری ح، علیزاده س، پرزگری الف. استخراج دانش از داده‌های بیماران دیابتی با استفاده از روش درخت تصمیم. مجله مدیریت سلامت. ۱۳۹۲، ۱۶ (۵۳): ۵۸-۷۲.
- ۱۴- محمودی ع، عسکری مقدم ر، صادقیان س. مدل پیش‌بینی بیماری عروق کرونر با استفاده از شبکه‌های عصبی و گزینش متغیر مبتنی بر درخت رگرسیون و طبقه‌بندی. مجله دانشگاه علوم پزشکی شهرکرد. ۱۳۹۲، ۵ (۱۵): ۴۷-۵۶.
- ۱۵- نبوتی الف، عزیزی الف، عباسی الف، وکیلی ارکی ح، زارعی ح، رضوی الف. کاربرد داده کاوی در پیش‌بینی مرگ بیماران سوختگی. مجله مدیریت اطلاعات سلامت. ۱۳۹۲، ۶ (۱۰): ۷۸۹-۷۹۹.
- 16- Hian ChyeKoh, Gerald Tan. Data Mining Applications in Healthcare. Journal of Healthcare Information Management. Vol.19, No.2.
- 17- <http://www.tamin.ir/News/Item>.

# Evaluation of Data Mining Applications in the Health System

Leila Gholamhosseini<sup>1</sup>, Mostafa Damroodi<sup>\*1</sup>

---

## Abstract

**Introduction:** Extensive amounts of data stored in medical databases require the development of specialized tools for accessing the data, data analysis, knowledge discovery, and the effective use of the data. Data mining is one of the most important methods. The article sketches the used Data Mining techniques, and illustrates their applicability to medical diagnostic and prognostic problems.

**Materials and Methods:** The current study were searched English and Persian databases including Magiran, Iranmedex, SID, Google Scholar, OVID, Scopus, and PubMed by using keywords such as "Data Mining", "Knowledge Discovery" and "Health". Related articles were published and assessed during 1998-2013.

**Findings:** Data mining is a science that is searched automatically in the large amount of data for finding models and association rules in them where other statistical analysis cannot do that. The medical science is one of sciences that need to use of these tools for analyzing the large amount of data and creating predictive model with the new computation ways. In medical sciences, discovery and early diagnosis of the diseases can restrict the fatal diseases such as cancer and they save a people's life. This research is shown that the data mining prediction provide necessary tools for the researcher and the physician to improve in the prevention of disease, diagnosis ways and their treatment programs.

**Discussion and Conclusion:** Nowadays, in the medical sciences, data collection of different diseases is very important. Recent development related to information technology & software has helped to have the better survey from producing massive data and could search the hidden knowledge in the data and create a new science by using different sciences including statistics, computers, and machine learning, and etc.

**Keywords:** data mining, health, knowledge discovery

---

---

1- (\*Corresponding author) Department of Health Information Technology, Faculty of Paramedicine, AjA University of Medical Sciences, Tehran, Iran. E-mail: mostafadamroodi1992@yahoo.com