

داده‌کاوی بالینی: مروری بر تکنیک‌های داده‌کاوی در دیابت

حامد مهدی زاده^{1*}، علیرضا برآنی²

چکیده

مقدمه: ارائه خدمات مراقبتی و بهداشتی افراد مبتلا به بیماری دیابت اطلاعات مفیدی ایجاد می‌کند که از این اطلاعات می‌توان برای شناسایی، درمان، مراقبت‌های بعدی و حتی پیشگیری از بیماری دیابت استفاده نمود. از طرفی کاوش و بررسی حجم زیادی از این اطلاعات، نیازمند استفاده از روش‌های موثر و کارآمدی برای یافتن الگوهای مربوط در این اطلاعات می‌باشد که استفاده از تکنیک‌های مختلف داده‌کاوی به‌خصوص دسته‌بندی و الگوهای تکرار شونده می‌تواند کمک شایانی در این زمینه باشد.

روش‌ها: پایگاه‌های اطلاعاتی Pubmed, Scopus, Google Scholar, Science Direct با هدف یافتن مقالات، جستجو و مقالات انگلیسی منتشر شده در سال‌های 2005 تا 2015 مورد بررسی قرار گرفتند. مقالات به‌دست آمده از نظر جمعیت مورد مطالعه، مجموعه داده‌های مورد استفاده و روش‌های داده‌کاوی ارزیابی شدند.

یافته‌ها: از میان 2144 مقاله به‌دست آمده در جستجوی اولیه، تعداد 38 مقاله مرتبط با موضوع مطالعه، انتخاب و مورد بررسی قرار گرفت. نتایج این بررسی نشان می‌دهد که الگوریتم‌های خوشه‌بندی، قوانین انجمنی و هوش مصنوعی از پرکاربردترین تکنیک‌های داده‌کاوی می‌باشند که برای تشخیص و پیش‌بینی احتمال ابتلا به بیماری دیابت با موفقیت مورد استفاده قرار گرفته‌اند.

نتیجه‌گیری: پیاده‌سازی روشی که بتواند امکان ابتلا یا عدم ابتلای افراد به دیابت را مشخص کند، گام مهمی در کنترل بیماری دیابت خواهد بود. با توجه به مطالعات انجام شده، داده‌کاوی می‌تواند به‌عنوان روشی موثر در پیشگیری، درمان و کشف ارتباط بین بیماری دیابت و عوامل خطر آن، موجب پیشرفت‌های قابل توجه در حوزه تحقیقات دیابت و ارائه مراقبت‌های بهداشتی بهتر برای این گروه از بیماران باشد.

واژگان کلیدی: داده‌کاوی بالینی، داده‌کاوی دیابت، دیابت

1- گروه انفورماتیک پزشکی و مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

2- گروه مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

***نشانی:** تهران، تجریش، میدان قدس، ابتدای خیابان دربند، دانشکده پیراپزشکی، کد پستی: 1971653313، تلفن: 02122718531، پست

الکترونیک: H.mehdizadeh@sbmu.ac.ir

تاریخ پذیرش: 1395/02/29

تاریخ درخواست اصلاح: 1395/02/20

تاریخ دریافت: 1394/12/03

مقدمه

بیماری در مراحل اولیه به دلیل پیچیدگی و وابستگی به عوامل درونی مختلف بسیار دشوار است [9]. بنابراین پیاده‌سازی روشی که بتواند امکان تشخیص صحیح ابتلا یا عدم ابتلا به دیابت را مشخص کند، می‌تواند گام مهمی در پیشگیری و کنترل این بیماری به‌خصوص در مراحل ابتدایی آن باشد. به‌همین دلیل متخصصان پزشکی نیاز به یک روش یا ابزار پیش‌بینی قابل اطمینان، برای تشخیص بیماری دیابت دارند [10]. داده‌کاوی یکی از این ابزارهای مفید است. هدف از این مطالعه بررسی تحقیقات انجام شده در حوزه بیماری دیابت می‌باشد که از تکنیک‌های مختلف داده‌کاوی برای پیش‌بینی، درمان یا مدیریت افراد مبتلا به دیابت استفاده کرده‌اند.

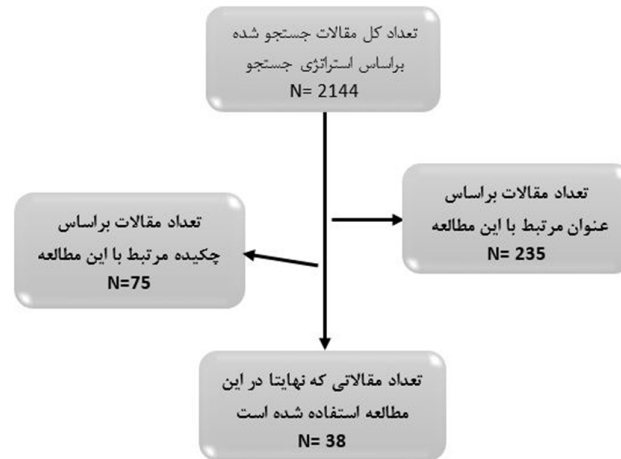
روش‌ها

مطالعه حاضر از نوع مروری نقلی بوده و پایگاه‌های اطلاعاتی Pubmed, Scopus, Google scholar, Web of Science, Ebsco, با هدف یافتن منابع مربوط جستجو گردیدند. این راهبرد جستجو محدود به منابع انگلیسی و در محدوده زمانی ده سال (2005-2015) بود. برای جستجو در بانک‌های اطلاعاتی انگلیسی از عملگرها و کلید واژه‌های زیر استفاده شد:

"clinical data mining" [All Fields] OR "data mining applications" [All Fields] OR "data mining" [All Fields] AND "diabetes" [All Fields]

نتیجه جستجوی کلی برای منابع مرتبط با داده‌کاوی در مراقبت بهداشتی 2144 بود که با محدود کردن نتایج جستجو در حوزه دیابت، 1861 منبع که خارج از حیطه موضوعی این مطالعه بودند حذف شدند. در مرحله بعدی با اعمال فیلتر محدوده زمانی تاریخ یکم آوریل 2005 تا 31 مارس 2015، دسترسی به متن کامل و زبان انگلیسی برای نتایج جستجو، این تعداد به 75 عدد کاهش یافت. در مرحله پایانی نیز با حذف مقالات تکراری و بررسی ارتباط مقالات با موضوع مورد مطالعه، 38 مقاله واجد شرایط برای این مطالعه انتخاب شدند. شکل یک مراحل مربوط به جستجو و انتخاب منابع را نشان می‌دهد.

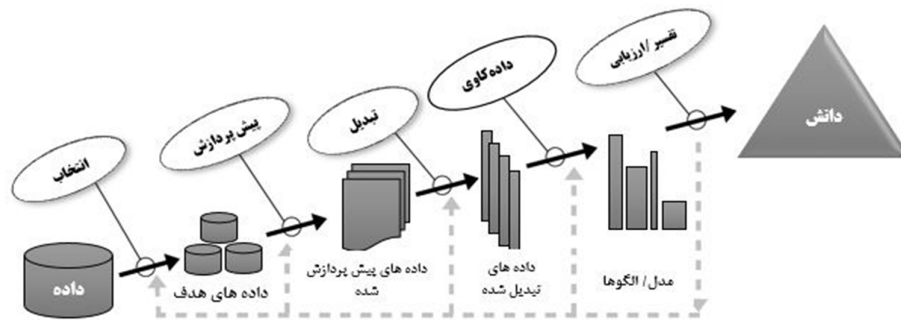
دیابت شامل گروهی از بیماری‌های متابولیک است که به دلیل عدم تولید کافی انسولین در بدن و عدم پاسخ‌دهی مناسب سلول‌ها به انسولین تولید شده، میزان قند خون فرد مبتلا افزایش می‌یابد [1]. بیماری دیابت یک مشکل جدی بهداشت جهانی و یکی از شایع‌ترین بیماری‌های عصر حاضر است. این بیماری معروف به بیماری جامعه مدرن است و میزان بروز این بیماری به‌صورت صعودی در حال افزایش است [2]. در حال حاضر 285 میلیون نفر در سراسر جهان مبتلا به دیابت هستند و پیش‌بینی می‌شود این میزان در سال 2030 به 438 میلیون نفر افزایش پیدا کند [3]. همچنین بنا بر گزارش اطلس جهانی دیابت (World Diabetes Atlas)، در هر هفت ثانیه یک فرد مبتلا به دیابت فوت می‌کند و تنها در سال 2014 چهارمیلیون و نهصد هزار نفر جان خود را بر اثر ابتلا به این بیماری از دست داده‌اند [4]. با وجود گستردگی و شیوع بالای این بیماری، تا کنون روشی برای درمان و از بین بردن این بیماری پیدا نشده است، هر چند روش‌هایی برای تشخیص و کنترل این بیماری وجود دارد [5]. دیابت بر دونوع می‌باشد، دیابت نوع یک وابسته به انسولین نامیده می‌شود که توانایی تولید انسولین در بدن کاهش یا قطع می‌شود [6]. در دیابت نوع دوم یا دیابت ملیتوس، به‌خاطر کافی نبودن انسولین ترشح شده توسط بدن، و یا عدم پاسخ‌دهی سلول‌ها به انسولین ترشح شده موجود، شخص مبتلا دارای میزان قند خون بالایی می‌باشد [7]. تقریباً 90 درصد از مبتلایان به دیابت از نوع دوم رنج می‌برند و شناسایی زودهنگام افراد در معرض خطر و بیماران مبتلا به این نوع دیابت، یکی از چالش‌های مهم در زمینه پزشکی دیابت است [3]. تحقیقات پزشکی نشان داده است که 80 درصد از عوارض مربوط به دیابت نوع دوم، در صورت تشخیص در مراحل اولیه، قابل پیشگیری می‌باشد [8]. با این وجود هنوز هم یکی از مشکلات اساسی در مدیریت بیماری دیابت عدم تشخیص به‌موقع و در مواردی حتی عدم تشخیص صحیح بیماری دیابت می‌باشد که به دلیل عدم انتخاب الگوی مناسب تشخیصی از سوی پزشک معالج می‌باشد. در واقع تشخیص این



شکل 1- مراحل انتخاب منابع مورد استفاده در این تحقیق

داده‌کاوی و کاربرد آن در دیابت: داده‌کاوی روشی برای کشف الگوهای پنهان و استخراج اطلاعات معنی‌دار از مجموعه داده‌های بزرگ می‌باشد [11]. در واقع داده‌کاوی بخشی از فرایند استخراج دانش (Knowledge Discovery) می‌باشد که هدف آن دستیابی به دانش نهفته در داده‌ها با کمترین (یا عدم) دخالت انسانی است [12]. این روش شامل الگوریتم‌هایی نظیر قوانین انجمنی (Association Rules)، مدل بیزی

(Bayesian Model)، درختان تصمیم (Decision Trees)، شبکه عصبی مصنوعی (Artificial Neural Network) برای کشف اطلاعات جدید است و از نرم افزارهای مختلفی همچون Weka، R و Rapid Minder استفاده می‌شود [3]. در شکل 2 مراحل اکتشاف دانش و داده‌کاوی نشان داده شده است. شکل دو مراحل مربوط به استخراج دانش از داده‌های خام را نشان می‌دهد.



شکل 2- فرآیند استخراج دانش یا کشف دانش از داده‌های خام

در واقع داده‌کاوی می‌تواند به‌عنوان یک ابزار با ارزش برای محققان بیماری دیابت باشد تا بدین وسیله دانش و الگوی پنهان در حجم وسیع اطلاعات مربوط به دیابت را استخراج کنند و به متخصصان پزشکی در تشخیص این بیماری کمک کنند [20]. بر همین اساس، مطالعات مختلفی در سال‌های اخیر با هدف بررسی و به‌کارگیری الگوریتم‌های پرکاربرد داده‌کاوی در بیماری دیابت صورت گرفته که در ادامه با معرفی هر یک از این روش‌ها و توضیح مختصری در مورد آن‌ها، نتایج مربوط به نمونه‌هایی از مطالعاتی که به‌صورت کاربردی از این روش‌ها استفاده کرده‌اند، ارائه می‌شود.

قوانین انجمنی²

قوانین انجمنی یکی از الگوریتم‌های ایده آل داده‌کاوی برای کشف و پیش‌بینی الگوهای جامع پنهان، در پایگاه داده‌های پزشکی می‌باشد. این الگوریتم بیشتر در مواردی مورد استفاده قرار می‌گیرد که هدف جستجو و کشف قوانین جدید باشد. در واقع این الگوریتم رابطه بین یک جفت (Y, X) از مجموعه‌ای از صفات، مشخص شده توسط $X \rightarrow Y$ ، که در آن X مقدم است و Y ناشی از $X \rightarrow Y$ می‌باشد را تعیین می‌کند. به عبارتی اگر X اتفاق بیافتد، سپس Y هم اتفاق می‌افتد [21].

این قوانین به‌صورت تفسیری بوده و در دو مرحله تولید می‌شود. در گام نخست مجموعه‌ای از آیت‌های مکرر یا الگوهای مکرر تولید شده و در گام دوم این الگوها برای تولید قوانین انجمنی استفاده می‌شوند. قوانین انجمنی به‌طور گسترده‌ای در کشف رابطه بین علامت و نوع سندرم استفاده می‌شوند و می‌توانند زیرمجموعه جمعیت در معرض خطر یک بیماری را تشخیص دهند [22]. البته این روش تنها برای پیش‌بینی به‌کار نمی‌رود و متمرکز بر مشخص کردن این است که احتمال ابتلا به یک حالت بیماری زمانی که افراد در معرض ترکیب متفاوتی از عوامل خطرزا قرار بگیرند چگونه خواهد بود [23].

برای نمونه، Rane و همکاران در 2013 از مجموعه داده‌های NIMARY متعلق به یک مرکز فوق تخصصی دیابت برای یافتن عوامل ایجاد کننده دیابت نوع دوم و روابط بین آن‌ها استفاده کردند. این مجموعه داده شامل 900 نمونه و حدود 30

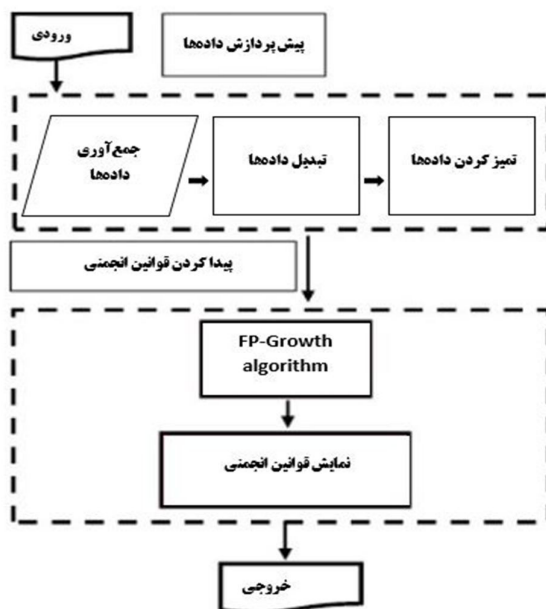
امروزه، تلاش‌های بسیاری با هدف به‌کارگیری داده‌کاوی در حوزه بهداشت و درمان صورت گرفته است و با نام داده‌کاوی بالینی شناخته می‌شود که با توسعه و پیشرفت ابزارهای فن آوری اطلاعات و ارتباطات می‌تواند راهکار نوین و مفیدی برای کشف الگوهای موجود در داده‌های سلامت باشد [13]. سیستم‌های مراقبت‌های بهداشتی مملو از اطلاعاتی است که دانش قابل توجهی در آن نهفته است بنابراین به تکنیک‌ها و ابزارهایی برای استخراج اطلاعات مفید از این مجموعه داده‌های بزرگ نیاز است [14]. اولین مزیت استفاده از این روش، امکان تحقیق برای کشف الگوهای پنهان در مجموعه داده‌های موجود در حوزه بهداشت و درمان می‌باشد که از این الگوها می‌توان برای تشخیص بیماری‌ها و همچنین درمان آن‌ها استفاده نمود [15].

در این میان بیماری دیابت به چند دلیل یکی از بیماری‌های مناسب برای کاربرد داده‌کاوی در پزشکی می‌باشد. دیابت از جمله بیماری‌هایی است که عوارض جانبی متعدد و دردناکی همچون نارسایی قلب و کلیه، نایبایی و حتی قطع عضو را در پی دارد که هر کدام از آن‌ها به تنهایی هزینه‌های مادی و معنوی زیادی را برای فرد و جامعه همراه دارد [16]. از طرفی به دلیل شیوع بالای این بیماری، شناسایی و کشف الگوهای بیماری و روش درمان آن می‌تواند موجب صرفه‌جویی قابل ملاحظه‌ای در زمان و هزینه‌های اقتصادی برای مؤسسات ارائه کننده سلامت و همچنین بیماران شود [17]. از طرفی به دلیل تاثیر عوامل متعدد در این بیماری، متغیرهای بسیاری در تصمیم‌گیری‌های پزشکی مربوط به آن در نظر گرفته می‌شوند [18]. بنابراین پیاده‌سازی روشی که بتواند امکان تشخیص صحیح ابتلا به دیابت و یا عدم ابتلا را مشخص کند می‌تواند گام مهمی در پیشگیری و کنترل این بیماری به‌خصوص در مراحل ابتدایی آن باشد [8]. که این موارد به همراه وجود بانک‌های اطلاعاتی متعددی که اطلاعات مربوط به بیماران مبتلا به دیابت را در دوره‌های زمانی متفاوت ثبت و نگهداری می‌کنند. مانند مجموعه داده‌های پیما¹ (داده‌های مربوط به سرخپوستان آمریکای شمالی)، موجب می‌شود بیماری دیابت به‌عنوان یکی از حوزه‌های مناسب برای به‌کارگیری داده‌کاوی باشد [19].

² Association Rule

¹ PIMA

شد. در این مطالعه از نرم‌افزار Weka استفاده شد. نتایج این مطالعه نشان داد که با استفاده از قوانین انجمنی می‌توان عوامل خطر مستقیم و غیرمستقیم ایجاد کننده دیابت نوع دوم را مشخص کرده و از این طریق به پزشکان برای جستجوی داده‌ها و فهم بهتر قوانین کشف شده در مورد بیماری دیابت کمک کند [10]. در شکل 3، ساختار مدل استفاده شده در این مطالعه نمایش داده شده است.



شکل 3- معماری مدل استفاده شده در مطالعه Rane و همکاران [10]

اعصاب، کلیه، چشم و عوارض گردش خون محیطی بودند پرداختند. نمونه‌ای از قوانین انجمنی استخراج شده توسط این مدل عبارت بودند از: (الف) میزان عوارض مربوط به بیماری دیابت نوع دوم در هر دو جنس زن و مرد یکسان است. (ب) زنان مبتلا به دیابت نوع دو که در رده سنی 50 تا 79 سال می‌باشند، بیشتر از سایرین مستعد ابتلا به عوارض کلیوی، اعصاب و چشم می‌باشند. (ج) افراد مبتلا به دیابت که در محیط‌ها کاری مثل کارخانه کار می‌کنند بیشتر از سایرین مستعد ابتلا به عوارض مربوط به چشم، مغز و اعصاب و عوارض گردش خون محیطی می‌باشند [1]. در این مطالعه با به‌کارگیری تکنیک داده‌کاوی برای بررسی داده‌های موجود، امکان مشخص کردن جمعیت در معرض خطر میسر شد. Chun و همکاران نیز در سال 2008 با استفاده از تکنیک قوانین انجمنی به مطالعه

ویژگی مربوط به هر نمونه در رابطه با دیابت نوع دوم بود. در این تحقیق به منظور افزایش کیفیت داده‌ها، بر پایه نظرات افراد متخصص از تکنیک باینینگ (Equal Interval Binning) برای پیش پردازش داده‌ها استفاده شد. در این روش ابتدا داده‌های عددی به داده‌های طبقه‌ای تبدیل و سپس از روش الگوهای متناوب افزایشی (Frequent Pattern Growth) برای شناسایی و تولید قوانین انجمنی قوی از مجموعه داده‌های بیماران استفاده

نمونه‌هایی از قوانین انجمنی استخراج شده در مطالعه Rane و همکاران [10] که امکان پیش‌بینی احتمال ابتلا به دیابت و علت آن را مشخص می‌کند:

Rule 1. {Urine Alb. < 300}{Heart problem is absent}{Ceratine is Negative}{TG < 250}{Uric Acid is Absent}{LDL is Low}-->{T2DM is present} 96.66%
Rule 2. {Ceratine is Negative}{TG < 250}{SGPT has no value}{HDL < 35}{LDL is Low}-->{T2DM is present} 96.66%

Rule 3. {Urine Alb. < 300}{Ceratine is Negative}{TG < 250}{SGPT has no value}{HDL < 35}{LDL is Low}-->{T2DM is present} 96.66%

Rule 4. {Heart problem is absent}{Ceratine is Negative}{TG < 250}{SGPT has no value}{HDL < 35}{LDL is Low}-->{T2DM is present} 96.66%

در سال 2012 Kurutach و همکاران در مطالعه‌ای که با هدف ارائه یک مدل برای کشف قوانین انجمنی انجام شد، به بررسی داده‌های مربوط به بیماران دیابتی نوع دوم که دارای نارسایی

ارتباط بین دیابت، سن، فشارخون و چربی بالا پرداختند. داده‌های مورد استفاده در این مطالعه مربوط به پایگاه داده بیمه بهداشت و درمان تایوان بود. نتایج این مطالعه نشان داد که شیوع دیابت با افزایش سن بیشتر شده و این عامل در زنان مسن شدیدتر است. همچنین یک رابطه قوی بین دیابت و بسیاری از بیماری‌هایی چون پوسیدگی دندان، پالپ (pulpits)، التهاب حاد لثه مشخص شد. این مطالعه نشان داد که با استفاده از تکنیک‌های داده‌کاوی می‌توان رابطه بین بیماری‌ها و عوامل دیگر با دیابت را مشخص کرد [24]. در واقع استفاده از قوانین انجمنی در داده‌کاوی دیابت، در بیشتر مطالعات انجام شده دارای نقش پیش‌بینی کننده برای مشخص کردن جمعیت در معرض خطر، عوامل ایجاد کننده و احتمال ابتلا به این بیماری می‌باشد.

الگوریتم‌های خوشه بندی¹

یکی دیگر از تکنیک‌های پرکاربرد داده‌کاوی، استفاده از الگوریتم‌های دسته‌بندی می‌باشد. فرآیند تجزیه و تحلیل و گروه‌بندی داده‌ها به کلاس‌های مختلف طوری که اشیاء در داخل یک کلاس شباهت بیشتری به یکدیگر دارند اما نسبت به کلاس دیگر بیش‌ترین تفاوت را دارند. هدف اصلی خوشه‌بندی در دیابت، پیدا کردن گروه‌های مختلف بیماران مبتلا با علائم مشابه در یک گروه، اما با نشانه‌های متفاوت از گروه‌های دیگر می‌باشد [19]. در ادامه به نمونه ای از مطالعات انجام شده در این زمینه اشاره می‌شود. در سال 2014 Diwani و همکاران با استفاده از نرم‌افزار Weka به بررسی و مقایسه عملکرد روش‌های مختلف داده‌کاوی در شناسایی بیماران دیابتی نوع دوم پرداختند. در این مطالعه روش‌های یادگیری ماشین نظارت شده مثل الگوریتم درخت تصمیم‌گیری و نایو بیز بر روی مجموعه داده‌های دیابت پیما اعمال و سپس عملکرد هر کدام از این الگوریتم‌ها و میزان موفقیت آن‌ها در شناسایی و تشخیص بیماران دیابتی مورد ارزیابی قرار گرفت. نتایج به دست آمده از این مطالعه نشان داد که الگوریتم نایو بیز با میزان دقت 76/30 درصد عملکرد بهتری نسبت به الگوریتم درخت تصمیم با میزان دقت 73/82 درصد داشته و بهینه‌تر می‌باشد [7].

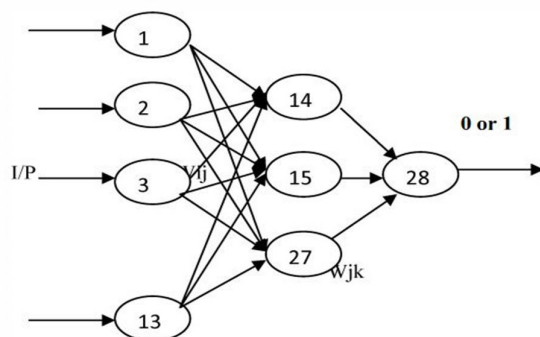
در مطالعه دیگری در سال 2013 Aljumah و همکاران با هدف تجزیه و تحلیل پیشگویانه درمان دیابت در عربستان انجام دادند، از تکنیک ماشین‌های بردار پشتیبان² برای کشف الگوهایی که بهترین درمان را برای بیماران دیابتی در سنین مختلف مشخص کند استفاده کردند. مجموعه داده‌های مورد استفاده در این تحقیق مربوط به مجموعه داده‌های مربوط به عوامل خطر بیماری‌های غیرواگیردار کشور عربستان در سازمان جهانی بهداشت (WHO) مورد تجزیه و تحلیل قرار گرفت. برای این منظور در این مطالعه از تکنیک آنالیز رگرسیون و ابزار ODM³ به‌عنوان یک نرم‌افزار داده‌کاوی برای آنالیز داده‌ها استفاده شد. در این مطالعه مجموعه داده‌های مورد نظر به‌منظور تعیین میزان اثر بخشی درمان‌های متفاوت، در دو گروه سنی جوان و پیر مورد بررسی قرار گرفت. نتایج این مطالعه نشان می‌داد که میزان اثربخشی درمان در دو گروه سنی متفاوت است [9]. در واقع در این مطالعه الگوی مناسبی برای یک طرح درمانی اثربخش ارائه شد که مشخص می‌کرد درمان دارویی در گروه سنی جوانان باید دیرتر شروع شود تا عوارض جانبی این بیماری کاهش یابد. و در مقابل درمان دارویی در گروه سنی افراد پیر باید بلافاصله همراه با درمان‌های دیگر شروع شود چرا که گزینه جانشین دیگری وجود ندارد. حالت مشترک برای درمان این بیماری در هر دو گروه، تجویز دارو برای بیماران مبتلا به دیابت نوع دوم به‌منظور کنترل عوارض دیابت بود. Aljarullah و همکاران نیز در سال 2011 با استفاده از مجموعه داده‌های پیما، مدلی برای پیش‌بینی و تشخیص بیماری دیابت نوع دوم ارائه کردند. قبل از پردازش داده‌ها، به‌منظور بهبود کیفیت داده‌ها از تکنیک‌های پیش پردازش داده استفاده شد. تکنیک‌های پیش‌پردازش برای شناسایی و انتخاب ویژگی‌ها، مدیریت داده‌های مفقود و داده‌های گسسته به‌کار می‌رود. در مرحله بعدی از درخت تصمیم‌گیری برای ساخت مدل استفاده شد. از مجموع 724 نمونه انتخاب شده، 566 نمونه به‌درستی و براساس ویژگی‌های انتخاب شده، توسط مدل طبقه بندی شده بودند. دقت مدل ارائه شده در تشخیص بیماری دیابت در این مطالعه 78 درصد بود که از نظر بالینی قابل ملاحظه به‌نظر می‌رسد [25]. در واقع الگوریتم‌های دسته‌بندی و خوشه‌بندی

² Support Vector Machine

³ The oracle Data Miner

¹ Clustering

مطالعه حدود 13 پارامتر مانند جنسیت، سن، وزن، قد، افزایش تشنگی، افزایش گرسنگی، افزایش اشتها، استفراغ و غیره مربوط به داده‌های 100 نفر از گروه‌های سنی، جنس و سبک زندگی مختلف، انجام شد. در این مطالعه از نرم افزار MATLAB استفاده شد. مدل شبکه عصبی ارائه شده در این مطالعه دارای 27 نود² بود که 13 نود ورودی و 13 نود لایه پنهان و یک نود خروجی داشت. 13 ویژگی مورد نیاز برای نودهای ورودی عبارت بودند از: سن، جنس، وزن، قد، کاهش وزن، افزایش گرسنگی، افزایش تشنگی، افزایش اشتها، حالت تهوع، استفراغ، خستگی، تاری دید و عفونت مثانه یا دستگاه تناسلی که بعد از ورود این مقادیر به مدل با محاسباتی که در لایه پنهان این مدل انجام می‌شد و در نود خروجی نیز مقدار صفر به معنی عدم ابتلا و یا مقدار یک به معنی ابتلا به بیماری دیابت نوع دوم بود. میزان دقت این مدل در تشخیص صحیح افراد مبتلا به دیابت 92/8 درصد بود. در صورتی که ویژگی‌ها و مقادیر مورد نیاز (مقادیر ورودی) به درستی وارد مدل شوند این مدل در تشخیص افراد مبتلا به دیابت نوع دوم مطمئن عمل می‌کند [27]. شکل 4 شمای مفهومی شبکه عصبی آموزش دیده شده برای تشخیص دیابت در این مطالعه را نشان می‌دهد.



شکل 4- شبکه عصبی آموزش دیده شده برای تشخیص دیابت

در بیشتر این مطالعات با هدف شناسایی و دسته‌بندی بیماران دیابتی و افراد مبتلا به کار رفته است.

سایر روش‌های داده‌کاوی در بیماری دیابت

علاوه بر روش‌های گفته شده بالا، مطالعات دیگری نیز وجود دارند که از الگوریتم‌های دیگر داده‌کاوی از جمله هوش مصنوعی برای تشخیص دیابت استفاده کرده‌اند. از آنجا که تشخیص بیماری دیابت از طریق تفسیر مناسب داده‌های دیابتی، یک مشکل مهم در دسته‌بندی است. روش‌های مختلف هوش مصنوعی برای حل این مشکل در دیابت استفاده می‌شوند.

Luo و همکاران در سال 2014، یک سیستم ارائه توصیه‌های درمانی برای دیابت نوع دوم ایجاد کردند که انواع داده‌های مختلف مربوط به زندگی افراد را از طریق سنسورهای پزشکی و دستگاه‌های همراه جمع‌آوری و سپس با تجزیه و تحلیل این داده‌ها از طریق الگوریتم‌های داده‌کاوی، در خروجی توصیه‌های فردی را برای هر بیمار به منظور کنترل سطح گلوکز خون ارائه می‌داد. در این مطالعه به منظور بررسی میزان اثربخشی بالینی سیستم ایجاد شده، به مدت سه ماه در محیط بالینی و بر روی بیماران مورد بررسی قرار گرفت. نتایج به دست آمده نشان داد که این سیستم می‌تواند یک ابزار نوین خود مراقبتی برای بیماران مبتلا به دیابت نوع دوم باشد و همچنین این سیستم با یکپارچه‌سازی داده‌کاوی، محاسبات همراه¹ و دانش پزشکی در یک سیستم هوشمند موبایل، می‌تواند برای افراد مبتلا به بیماری‌های مزمن مانند دیابت مفید باشد [26]. در این مطالعه به دلیل پیچیدگی زیاد و هزینه‌های بالای انجام آزمایش در محیط بالینی و بر روی انسان، آزمایش فوق بر روی یک گروه کوچک بیماران انجام شد بنابراین شواهدی مبنی بر اینکه در محیط واقعی بالینی نیز نتایج گفته شده مطابقت داشته باشد وجود ندارد. اما به طور کلی به عنوان سیستمی که قادر به پردازش داده‌های جمع‌آوری شده از سنسورهای پزشکی و ارائه توصیه‌هایی برای بیماران بود، به نوعی مزیت برای آنان محسوب می‌شود.

در مطالعه دیگری که توسط Sonukumari و همکاران در سال 2013 انجام شد، روشی هوشمند و موثر براساس شبکه‌های عصبی برای تشخیص خودکار دیابت پیشنهاد شد. در این

² Node

¹ Mobile computing

Ramezankhani و همکاران [23] انجام شد. در این مطالعه از بانک اطلاعاتی مطالعه قند و لیپید تهران استفاده شد. جمعیت مورد مطالعه در این تحقیق حدود 6647 نفر بودند که همگی زیر بیست سال سن داشته و طی دوازده سال به‌منظور تجزیه و تحلیل الگوهای پرخطر ابتلا به بیماری دیابت، پیگیری شدند. ویژگی‌هایی که در این مطالعه مورد بررسی قرار گرفتند شامل مشخصات جمعیت‌شناختی و تن سنجی، استعمال دخانیات، سابقه پزشکی و سابقه مصرف دارو و اقدامات آزمایشگاهی بود. نتایج این مطالعه نشان داد که مهم‌ترین عامل خطر بروز دیابت در زنان، چاقی و بعد از آن سابقه خانوادگی ابتلا به دیابت می‌باشد. در مورد مردان نیز، چاقی، کلسترول بالای خون، فعالیت بدنی کم و بیماری مزمن کلیوی به‌عنوان عوامل خطر برای ابتلا به دیابت اعلام شد. این مطالعه نشان داد که تکنیک قوانین انجمنی، روش مناسبی برای تعیین فراوانی رخداد ترکیبی از متغیرها و ویژگی‌های افرادی که مبتلا به دیابت هستند، می‌باشد.

در مطالعه دیگری، Habibi و همکاران در سال 2015 به بررسی یک مدل پیش‌بینی کننده با استفاده از عوامل خطر مربوط به دیابت نوع دوم پرداختند. در این مطالعه از داده‌های مربوط به پایگاه داده سیستم کنترل دیابت تبریز استفاده شد که مربوط به غربال‌گری انجام شده در فاصله زمانی 2009 تا 2011 بود. در این مطالعه با استفاده از درخت تصمیم‌گیری مدلی برای غربال‌گری بیماران دیابتی بدون نیاز به انجام تست‌های آزمایشگاهی ارائه شد. دقت این مدل در شناسایی افراد دیابتی 0/71 (1 به معنی تشخیص 100 درصد) بود [29]. جدول 5 به مطالعات مختلف انجام شده در حوزه داده‌کاوی دیابت را به همراه الگوریتم‌ها و نرم‌افزارهای مورد استفاده نشان می‌دهد.

در سال 2012 نیز Marciano و همکاران از یک پرسپترون چندلایه (Multi Layer Perceptron) به‌عنوان یک روش داده‌کاوی برای تشخیص بیماری دیابت بر روی داده‌های پیمای استفاده کردند. میزان دقت مدل به‌دست آمده با روش‌های دیگری از جمله درخت تصمیم‌گیری و دسته بندهای بیزین مورد بررسی قرار گرفت. در این مطالعه برای اعتبارسنجی، هر کدام از الگوریتم‌ها ابتدا آموزش و مدل دسته بند آن ساخته شد و سپس مورد ارزیابی قرار گرفت. نتایج به‌دست آمده به‌طور کلی نشان داد که الگوریتم هوش مصنوعی به‌کار رفته در این مطالعه با میزان دقت 89/53 درصد، حساسیت (Sensitivity) 97/1 درصد و ویژگی (Specificity) 75/6 درصد، به میزان قابل توجهی نسبت به الگوریتم درخت تصمیم‌گیری و دسته بند بیزین قوی‌تر و بهینه‌تر می‌باشد [8].

مطالعات انجام شده در ایران

علاوه بر مطالعاتی که در حوزه داده‌کاوی دیابت در کشورهای مختلف انجام شده است، مطالعاتی نیز در ایران انجام شد. در سال 2014 Sheikhpour و همکاران مطالعه‌ای با هدف ارائه الگوریتمی برای پیش‌بینی بیماری دیابت انجام دادند. در این مطالعه یک الگوریتم جدید که ترکیبی از روش کاهش بعد و الگوریتم کای نزدیک‌ترین همسایه (K-Nearest Neighbour) بود ارائه و با اعمال این الگوریتم بر روی 786 نمونه دیابتی و سالم از مجموعه داده‌های پیمای، مشخص شد که دقت این مدل در پیش‌بینی بیماری دیابت 82/09 درصد بود. نتایج این مطالعه نشان داد که کاهش ابعاد داده به‌همراه طبقه‌بندی داده‌ها می‌تواند در تشخیص بیماری دیابت برای پزشکان بسیار مفید باشد [28]. مطالعه دیگری نیز در همین سال و با هدف شناسایی عوامل خطر بروز دیابت نوع دوم با استفاده از قوانین انجمنی توسط

جدول 5- مطالعات مختلف انجام شده در حوزه داده‌کاوی دیابت

سال	محققین	هدف	مجموعه داده	روش داده‌کاوی	نرم افزار مورد استفاده	نتایج
2015	Ramezankhani [23]	شناسایی عوامل خطر بروز بانک اطلاعاتی قند و دیابت	لیبید تهران	Association Rule Mining	-	مهم‌ترین عوامل خطر: چاقی و سابقه خانوادگی ابتلا به دیابت
2014	[7] Diwani	مقایسه روش‌های تشخیص دیابت	PIMA	Decision Tree, J48/ Nave Baysean	Weka	اثبات بهینه بودن عملکرد الگوریتم نایو بیز نسبت به درخت تصمیم
2013	[9] Aljumah	کشف الگوی درمان متناسب با سن افراد	مجموعه داده ملی عربستان	Predictive Regression Analysis	ODM	شروع دیرتر درمان دارویی در جوانان برای کاهش عوارض جانبی
2013	[10] Rane	ارتباط عوامل ایجاد دیابت	NIMARY	Association/ FP-growth	-	شناسایی عوامل مستقیم و غیرمستقیم خطرزا در دیابت نوع دوم
2013	(30) Kumari	تشخیص هوشمند دیابت	-	Artificial Neural Networks	-	تشخیص افراد مبتلا به دیابت با دقت 92/8 درصد
2012	Marcano-Cedeno [8]	تشخیص هوشمند دیابت	PIMA	Multi Layer Perceptron	Weka	الگوریتم MLP بهینه‌تر از الگوریتم درخت تصمیم و بیزین
2012	[1] Kurutach	ارائه مدل برای پیش‌بینی دیابت	-	Association Rules	Weka	یکسان بودن عوارض بیماری دیابت نوع دوم در هر دو جنس
2012	[31] Karegowda	ارائه مدل برای دسته بندی (تعین ابتلا و عدم ابتلا)	PIMA	Decision Tree / k-means	-	دقت مدل ترکیبی از درخت تصمیم بتن‌هایی بیشتر است
2011	[25] AlJarullah	پیش‌بینی ابتلا به دیابت	PIMA	Decision Tree J.48	Weka	ارائه مدل تصمیم‌گیری با دقت 78 درصد
2010	[32] Nuwangi	رابطه دیابت و بیماری‌های دیگر	-	Association Rules	-	علامت‌های خاص سینه عامل خطر افزایش ابتلا به دیابت
2010	[33] Barakat	کشف الگو برای پیش‌بینی دیابت	-	Support Vector Machine	-	افزایش احتمال ابتلا به دیابت با افزایش میزان گلوکز از 200 میلی‌گرم،
2008	[24] Chan	ارتباط دیابت و سندرم متابولیک	پایگاه داده بهداشت ملی تایوان	Association Rules/ Apriori	-	رابطه قوی بین دیابت و پوسیدگی دندان و التهاب حاد لثه
2008	[34] Harle	طبقه‌بندی بصری و تفسیر داده	مجموعه داده انجمن دیابت آمریکا	Data Visualization	-	ارائه مدلی برای کمک به تصمیم‌گیری بالینی پزشک
2007	[35] Juan	مدل برای تشخیص دیابت	PIMA	Decision Tree/ C.45	-	میزان دقت مدل برای تشخیص بیماران دیابتی 92/05 درصد
2005	[36] Denzinger	کاوش داده‌های پزشکی دیابت	-	-	-	ارائه مدل داده‌کاوی مشارکتی

بحث

بهداشت و درمان، مورد توجه قرار گرفته است و مزایای متعددی از قبیل تشخیص علل بیماری و شناسایی روش درمان پزشکی و به‌دنبال آن کاهش هزینه‌ها دارد [32]. در طی سال‌های اخیر، تحقیقات پزشکی مختلفی در حوزه دیابت، از تکنیک‌های داده‌کاوی مختلفی استفاده کرده‌اند که به‌طور کلی می‌توان این تکنیک‌ها را به دو دسته تقسیم کرد، دسته اول که به منظور پیش‌بینی عوامل ایجاد کننده و احتمال وقوع بیماری و دسته دوم نیز با هدف تشخیص افراد مبتلا به این بیماری می‌باشند. تکنیک‌های به‌کار گرفته شده در دسته دوم، متمرکز بر جستجوی اطلاعات برای

اطلاعات تولید شده توسط سازمان‌های بهداشتی بسیار گسترده و پیچیده است و به‌همین دلیل تجزیه و تحلیل داده‌ها به‌منظور تصمیم‌گیری در مورد سلامت بیمار دشوار است [14]. بنابراین برای تجزیه و تحلیل و استخراج اطلاعات مهم از این داده‌های پیچیده، نیازمند ایجاد و استفاده از یک ابزار قدرتمند می‌باشد [37]. در حال حاضر، استفاده از داده‌کاوی به‌دلیل نیاز به یک روش تحلیلی مؤثر برای شناسایی اطلاعات ناشناخته و ارزشمند در صنعت

دسته ورودی بررسی می‌شود تا خروجی شبکه به خروجی مورد انتظار نزدیک‌تر شود. بدین ترتیب مدل به دست آمده می‌تواند برای پیش‌بینی یا تشخیص بیماری به کار گرفته شود [11]. براساس نتایج مطالعات ارائه شده، استفاده از این روش‌ها با هدف کشف ارتباط بین بیماری دیابت و عوامل خطر، پیش‌بینی و درمان آن، می‌تواند موجب پیشرفت‌های قابل توجه در حوزه تحقیقات دیابت و ارائه مراقبت‌های بهداشتی بهتر برای این بیماران باشد. هرچند بسیاری از قوانین و قاعده‌ها در داده‌کاوی بر روی بهبود کارایی فرایند کاوش یا استخراج متمرکز هستند، اما در حوزه تحقیقات و علوم پزشکی، میزان کارآمدی و موثر بودن فرایند کاوش مهم‌ترین عامل نبوده بلکه نیاز به کشف قوانین انجمنی مفیدتر و حذف قوانین غیر مفید می‌باشد. در واقع کشف دانش از پایگاه داده‌های پزشکی به منظور ارائه تشخیص‌های پزشکی بسیار مهم می‌باشد [10].

نتیجه‌گیری

یکی از مشکلات اساسی در مدیریت بیماری دیابت عدم انتخاب الگوی مناسب تشخیصی از سوی پزشک و در نتیجه عدم تشخیص به موقع و صحیح بیماری دیابت می‌باشد. براساس نتایج ارائه شده در مطالعات نقل شده در این مقاله، به نظر می‌رسد استفاده از الگوریتم‌های داده‌کاوی برای تفسیر داده‌ها با هدف پیش‌بینی و تشخیص بیماری دیابت، منجر به افزایش موفقیت پزشکان در شناسایی افراد مبتلا به این بیماری و تعیین جمعیت‌های در معرض خطر بیماری شده و به دنبال آن با انجام اقدامات درمانی به موقع می‌توان مانع از پیشرفت بیماری و یا افزایش عوارض و مشکلات جانبی این بیماری مزمن در زندگی روزمره این گروه از بیماران شد. به‌طور کلی به‌کارگیری دانش به دست آمده از اعمال تکنیک‌های داده‌کاوی بر روی مجموعه داده‌های بهداشتی مربوط به جامعه می‌تواند یک عامل کلیدی در حرکت به سمت بهبود سیستماتیک و گسترده در بهداشت و درمان باشد.

یافتن الگوهای مفید و قابل توجه در رابطه با بیماری دیابت و ارتباط این عوامل در ایجاد این بیماری می‌باشد [15]. بر همین اساس روش‌های مختلف داده‌کاوی برای پیش‌بینی، تشخیص و مدیریت بیماری دیابت به کار گرفته شده‌اند که سه گروه الگوریتم‌های قوانین انجمنی، خوشه‌بندی و شبکه عصبی در مطالعات بررسی شده در این تحقیق جزء مواردی بودند که بیش از سایر الگوریتم‌های داده‌کاوی مورد استفاده قرار گرفته‌اند.

قوانین انجمنی برای شناسایی روابط بین مجموعه‌ای از اقلام داده‌ای در یک بانک اطلاعاتی به کار می‌رود. این روش در زمانی که به دنبال کشف قوانین جدید هستیم مناسب می‌باشد [38]. در واقع با استفاده از این روش و استخراج قوانین انجمنی، منجر به کشف علل احتمالی بیماری خاص مثل دیابت در قالب قوانین انجمنی می‌شود. برای به کار گیری در حوزه پزشکی، قوانین انجمنی می‌توانند جمعیت در معرض خطر یک بیماری را تشخیص دهند [21]. این قوانین به صورت تفسیری بوده و رابطه بین عوامل خطر و ایجاد یک بیماری را در قالب قوانین مشخص بیان می‌کنند [23]. از طرفی استفاده از الگوریتم‌های خوشه‌بندی نیز برای مشخص کردن ابتلا یا عدم ابتلای افراد به دیابت صورت می‌گیرد. این الگوریتم‌ها از طریق تجزیه و تحلیل و گروه‌بندی داده‌های موجود، افرادی که دارای ویژگی‌ها یا نشانه‌های مشابه باشند در گروه بیمار و افرادی که فاقد آن نشانه‌ها باشند به عنوان افراد سالم تعیین می‌شوند [17]. در واقع هدف اصلی الگوریتم‌های خوشه‌بندی در دیابت، پیدا کردن گروه‌بندی موارد با علائم مشابه در یک گروه می‌باشد که به معنی مشخص شدن افراد مبتلا و سالم از یکدیگر می‌باشد و می‌تواند به متخصصان پزشکی در تشخیص این بیماری کمک کند [34]. دسته سوم نیز مربوط به شبکه‌های عصبی مصنوعی می‌باشد که برای دسته‌بندی و تشخیص الگوها در بین داده‌ها به کار می‌روند. این الگوریتم‌ها برای تولید خروجی‌های بهینه با داده‌های آزمایشی به صورت متناوب و افزایشی آموزش داده می‌شوند و هر بار نتایج مربوط به هر

مآخذ

1. Kurutach W, Kasemthaweesab P. Association Analysis of Diabetes Mellitus (DM) With Complication States Based on Association Rules. *7th IEEE Conference on Industrial Electronics and Applications (ICIEA)*; 2012; Singapore: IEEE.
2. Priya S, Rajalaxmi R. An Improved Data Mining Model to Predict the Occurrence of Type-2 Diabetes using Neural Network. *International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C); International Journal of Computer Applications (IJCA)* 2012.
3. Shivakumar B, Alby S. A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes. *International Conference on Intelligent Computing Applications*; 2014: IEEE.
4. Thirumal P, Nagarajan N. Applying Average K Nearest Neighbour Algorithm to Detect Type-2 Diabetes. *Australian Journal of Basic and Applied Sciences* 2014; 8 (7):128-34.
5. *IDF World Diabetes Atlas: International Diabetes Federation*; 2015 [cited 2015 9/13]. Available from: <https://www.idf.org/diabetesatlas>.
6. Patil B, Joshi R, Toshniwal D. Association rule for classification of type -2 diabetic patients. *Second International Conference on Machine Learning and Computing: IEEE Computer Society*, 2010.
7. Diwani S, Sam A. Diabetes Forecasting Using Supervised Learning Techniques. *ACSII Advances in Computer Science: an International Journal* 2014; 3(5):10-18.
8. Marcano-Cedeno A, Andina D. Data mining for the diagnosis of type 2 diabetes. *World Automation Congress (WAC)*; 2012; Puerto Vallarta, Mexico: IEEE.
9. Aljumah A, Siddiqui M. Application of data mining: Diabetes health care in young and old patients. *Computer and Information Sciences* 2013; 25:127-36.
10. Rane N, Rao M. Association Rule Mining on Type 2 Diabetes using FP-growth association rule. *International Journal of Engineering and Computer Science* 2013; 2(8):2481-85.
11. Tomar D, Agarwal S. A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology* 2013; 5 (5):241-66.
12. Jacob G, Ramani R. Data Mining in Clinical Data Sets: A Review. *International Journal of Applied Information Systems (IJ AIS)* 2012.4(6).15-26
13. Iavindrasana J, Cohen G, Depeursinge A, Müller A et al. *Clinical Data Mining: a Review*. In: Geissbuhler A, editor. *IMIA Yearbook of Medical Informatics* 2009.
14. Shipeng Yu, Rao B. Introduction to the Special Section on Clinical Data Mining. *ACM SIGKDD Explorations Newsletter* 2012; 14 (1).
15. Pradhan M. Data Mining and Health Care: Techniques of Application. *Journal of Engineering and Computer science* 2014; 1 (2):18-26.
16. Rexeena X, Devi S, Saranya S. Risk Assessment for Diabetes Mellitus using Association Rule Mining. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 2014; 3 (2).324-8.
17. Rajesh K, Sangeetha V. Application of Data Mining Methods and Techniques for Diabetes Diagnosis. *International Journal of Engineering and Innovative Technology (IJEIT)* 2012; 2(3):224-29.
18. Bellazzi R, Abu-Hanna A. Data Mining Technologies for Blood Glucose and Diabetes Management. *Journal of Diabetes Science and Technology* 2009; 3 (3):603-112.
19. Iyer A, Sumbaly J, Sumbaly R. DIAGNOSIS OF DIABETES USING CLASSIFICATION MINING TECHNIQUES. *International Journal of Data Mining & Knowledge Management Process (IJD KP)* 2015; 5 (1):1-14.
20. Wasan S, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics. *Data Science Journal* 2006; 5:119-26.
21. Ordonez C, Zhao K. Evaluating association rules and decision trees to predict multiple target attributes. *Intelligent Data Analysis* 2011; 15(2):173-92.
22. Lekha A, Srikrishna C, Vinod V, editors. Utility of association rule mining: A case study using Weka tool. *International Conference on Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT)*; 2013; Tiruvannamalai: IEEE.
23. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F. An Application of Association Rule Mining to Extract Risk Pattern for Type 2 Diabetes Using Tehran Lipid and Glucose Study Database. *Int J Endocrinal Metab* 2015; 13(2).
24. Chan C. Discovery of Association Rules in Metabolic Syndrome Related Diseases. *International Joint Conference on Neural Networks, IJCNN*, 2008; Hong Kong: IEEE. .
25. AlJarullah A. Decision Tree Discovery for the Diagnosis of Type II Diabetes. *International Conference on Innovations in Information Technology*; 2011; Abu Dhabi: IEEE.
26. Luo Y, Ling C, Schuurman J, Petrella R. GlucoGuide: An Intelligent Type-2 Diabetes Solution Using Data Mining and Mobile Computing. *IEEE International Conference on Data Mining Workshop*; 2014; Shenzhen: IEEE

27. Sonukumari A. A data mining approach for the Diagnosis of Diabetes Mellitus. *7th International Conference on Intelligent Systems and Control*, 2013; Coimbatore, Tamil Nadu, India.
28. Sheikhpour R, AghaSarram M. Diagnosis of Diabetes Using an Intelligent Approach Based on Bi-Level Dimensionality Reduction and Classification Algorithms. *Iranian Journal of Diabetesband Onesity* 2014; 6(2):74-84.
29. Habibi S, Ahmadi M, Alizadeh S. Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining. *Global Journal of Health Science* 2015; 7(5):304-10.
30. Kumari S, Singh A. A Data Mining Approach for the Diagnosis of Diabetes Mellitus. *71th International Conference on Intelligent Systems and Control (ISCO)*; 2013: IEEE.
31. Karegowda A, Gowda A, Punya V, Jayaram M, Manjunath A. Rule based Classification for Diabetic Patients using Cascaded K-Means and Decision Tree C4. 5. *International Journal of Computer Applications* 2012; 45(12):45. 35.
32. Nandita R, Rao M. Association Rule Mining on Type 2 Diabetes using FP-growth association rule. *International Journal of Engineering And Computer Science* 2013; 2(8):2481-85.
33. Barakat N, Bradley A, Barakat M. Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus. *IEEE Transactions on information technology in biomedicine*, 2010; 14(4):1114-20.
34. Harle C, Neill D, Padman R. An information visualization approach to classification and assessment of diabetes risk in primary care. *3rd Inform Workshop on Data Mining and Health Informatics* 2008.
35. Juan G, Sen-lin L, Hong-bol J, Tie-mei Z, Yi-wen H. Type 2 diabetes data processing with EM and C4.5 algorithm. *IEEE/ICME International Conference on Complex Medical En2ineerin2*; 2007; Beijing: IEEE.
36. Denzinger J, James R. CoLe: A Cooperative Data Mining Approach and Its Application to early diabetes detection. *Fifth IEEE International Conference on Data Mining (ICDM'05)*; 2005: IEEE.
37. Anjana R, Ali M, Pradeepa R, Deepa M et al. The need for obtaining accurate nationwide estimates of diabetes prevalence in India- rationale for a national study on diabetes. *Indian J Med Res* 2011; 133(4):369-80.
38. Nandita R, Rao M. Association Rule Mining on Type 2 Diabetes using FP-growth association rule. *International Journal of Engineering And Computer Science* 2013; 2(8):2481-85.

CLINICAL DATA MINING: AN OVERVIEW OF DATA MINING TECHNIQUES IN DIABETES

Hamed Mehdizadeh^{*1}, Alireza Baraani²

1. Department of Medical Informatics and Health Information Management ,School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2. Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ABSTRACT

Background: Provide a health care service to the patients with diabetes provides useful information that could be used to identify, treatment, following up and prevention of diabetes. Explore and investigation of large volumes of data requires effective and efficient methods for finding hiding patterns in the data. The use of various techniques of data mining in particular Classification and Frequent patterns can be helpful.

Methods: This article is a narrative review. We searched keywords related to application of data mining in the field of diabetes, through related databases, in English language articles published from 2005 to 2015. Also related articles in the selected articles list have been analyzed.

Results: From the 2144 articles obtained in the initial search, 38 articles related to the subject of study, were selected. Several studies shown that classification and clustering algorithms, association rules and artificial intelligence are the most widely used data mining techniques for predict the risk of diabetes has been successfully used.

Conclusion: The important step in control of diabetes, use of the methods that could determine the possibility or lack of diabetes. According to studies conducted in this area seem to use data mining techniques to prevent, treat and discover the connection between diabetes and its risk factors, can lead to significant improvements in the field of diabetes research and provide better health care for this group of patients.

Keywords: Clinical Data mining, Diabetes data mining, Diabetes

* School of Paramedical Sciences .Darband St. Quds Sq. Tajrish, Tehran, Postal code: 1971653313. Tel: 021-22718531, Email: H.mehdizadeh@sbmu.ac.ir