

مروری بر نقش و عملکرد رویکردهای هوش مصنوعی و داده کاوی در تشخیص بیماری دیابت

محمد وندجلیلی

کارشناس ارشد، گروه فناوری اطلاعات و ارتباطات دانشگاه مراغه، دانشگاه مراغه، مراغه، ایران

m.vandjalili@maragheh.ac.ir

چکیده

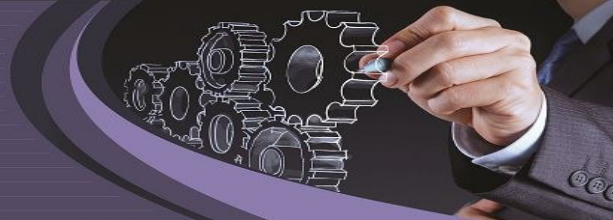
دیابت یکی از بیماری‌های مزمن و تاثیرگذار در جامعه جهانی بخصوص ایران می‌باشد، شناسایی و تشخیص دیابت در مراحل اولیه بیماری می‌تواند در درمان زود هنگام و سریع آن موثر بوده و عوارض جانبی و خطرناک آن شامل قطع عضو یا کوری جلوگیری به عمل آورد. تاکنون تشخیص دیابت عموماً بصورت آزمایشگاهی و سنجش عوامل بیماری (میزان پیشرفت آن) برپایه مشاهدات تجربی حاصل می‌گردد. امروزه با پیشرفت تکنولوژی‌های کامپیوتری به ویژه هوش مصنوعی و داده کاوی توانسته است تا رویکردی نوین را در حیطه تشخیص و طبقه‌بندی این بیماری مزمن ارائه نماید. در مطالعه حاضر ضمن بیان مختصری بر بیماری دیابت، رویکردهای مبنا در داده کاوی و عملکرد هوش مصنوعی در تشخیص این بیماری بصورت موردی مورد ارزیابی قرار گرفته است.

واژگان کلیدی: هوش مصنوعی، داده‌کاوی، تشخیص بیماری، دیابت.

مقدمه

دیابت یا بیماری قند یک نوع بیماری متابولیکی و مزمن بوده که نشانه آن قند خون بالاتر از ۱۲۰ (میلی‌گرم در دسی‌لیتر) در یک برهه طولانی است. این بیماری سه نوع اصلی نوع ۱، نوع ۲ و بارداری دارد؛ در دیابت نوع ۱ به علت تخریب سلول‌های بتای پانکراس توانایی تولید هورمون انسولین در بدن از بین می‌رود، این نوع از دیابت می‌تواند ناشی از یک بیماری خود ایمنی باشد و به دلیل نبود انسولین در بدن فرد مبتلا نیاز حتمی به تزریق انسولین وجود دارد این نوع از دیابت با نام‌های «دیابت وابسته به انسولین» یا «دیابت نوجوانان» نیز شناخته می‌شود. در دیابت نوع ۲، سلول‌های بدن به وجود انسولین پاسخ صحیحی نمی‌دهند و انسولین تولیدی نمی‌تواند عملکرد طبیعی خود را انجام داده و قند موجود در خون را کاهش دهد. این اتفاق در نهایت ممکن است به تخریب سلول‌های بتای پانکراس و توقف کامل تولید انسولین منجر شود. در ابتدای فرد به دیابت نوع دو عوامل ژنتیکی، چاقی و کم‌تحرکی نقش مهمی ایفا می‌کنند. نوع سوم آن دیابت زنانه یا دیابت بارداری نام دارد؛ این بیماری زمانی رخ می‌دهد که زنان باردار بدون سابقه دیابت دچار سطح بالای قند خون شوند (Nandita and Rao, 2013). هورمون انسولین و گلوکاگون تنظیم کنندگان قند خون می‌باشند، نقش گلوکاگون بالا بردن قند خون و نقش انسولین پایین آوردن قند خون توسط سازوکارهای مختلف است. در دیابت، سرعت و

هفتمین کنفرانس بین المللی تحقیقات بین رشته‌ای در مهندسی برق، کامپیوتر، مکانیک و مکترونیک در ایران و جهان اسلام



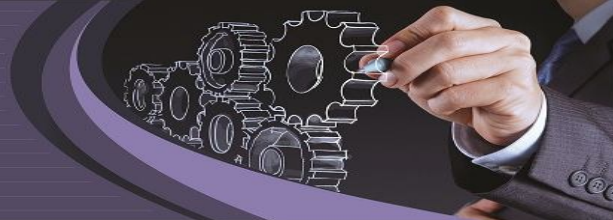
توانایی بدن در استفاده و سوخت و ساز کامل گلوکز کاهش می‌یابد از این رو میزان قند خون افزایش یافته که به آن هیپرگلیسمی می‌گویند. وقتی این افزایش قند در دراز مدت در بدن وجود داشته باشد، سبب تخریب رگ‌های بسیار ریز در بدن می‌شود که می‌تواند اعضای مختلف بدن همچون کلیه، چشم و اعصاب را درگیر کند. همچنین دیابت با افزایش ریسک بیماری‌های قلبی عروقی ارتباط مستقیمی دارد؛ لذا غربالگری و تشخیص زودرس این بیماری در افراد دارای بیماری‌های زمینه‌ای می‌تواند در پیشگیری از این عوارض مؤثر باشد. تشخیص و همچنین غربالگری دیابت با انجام آزمایش قند خون میسر است (Sheikhpour and AghaSarram, 2014).

طبق آمار بهداشت جهانی شیوع دیابت در بزرگسالان جهان $\frac{8}{8}$ است، که از این تعداد حدود 90٪ آن دیابت نوع 2 می‌باشد؛ همچنین نرخ شیوع جنسیتی آن به‌طور متوسط در زنان و مردان یکسان است. بیماری دیابت انواع مختلفی دارد که معمولاً در هنگام تشخیص متمایز می‌شوند؛ بنابراین تعیین نوع دیابت وابسته به شرایطی است که بیماری از خود نشان می‌دهد. خیلی از بیماران به راحتی در یکی از دسته‌های مشخص دیابت جای نمی‌گیرند، مثلاً فردی که دچار دیابت بارداری شده ممکن است بعد از پایان بارداری کماکان مبتلا به دیابت باقی بماند و لذا نوع دیابتش به نوع 2 تغییر داده شود؛ بنابراین برای پزشک متخصص نوع بیماری دیابت فرد نسبت به درک علت و سبب بیماری او و درمان مؤثر آن اهمیت بسیار کمتری دارد (Abdollahi and Nouri-Moghaddam, 2022). آخرین تقسیم‌بندی رسمی دیابت توسط انجمن دیابت آمریکا در 1997 بود و با وجود مشکلاتی که دارد هنوز توسط مراجع اصلی، مورد تأکید و استفاده است. با این وجود مدت‌های مدیدی است که مشخص شده دیابت یک بیماری هتروژن است و شاید نیاز به تقسیم‌بندی جدیدی باشد (Kavakiotis et al., 2017). جدیدترین دسته‌بندی پنج خوشه‌ای برای دیابت صورت گرفته که در این تقسیم‌بندی دیابت به جای دو نوع یک و دو در تقسیم‌بندی انجمن دیابت آمریکا به پنج خوشه به عنوان زیرگروه‌های دیابت دسته‌بندی شده‌اند. هدف این است که بتوان برای هر خوشه درمان مناسب را در زمان مناسب شروع کرد تا به نتایج بهتری نایل شویم. در 2 مارس 2018 گروهی از پژوهشگران در سوئد و فنلاند اعلام کردند که بر اساس تصویر پیچیده‌تری که آن‌ها از دیابت کشف کرده‌اند، دوره جدیدی در تجویز داروها برای دیابت آغاز شده‌است. این دانشمندان می‌گویند بیماری دیابت دارای پنج نوع مختلف است که درمان متناسب برای هر نوع می‌تواند جداگانه تهیه و تجویز شود. آن‌ها می‌گویند این نتیجه‌گیری آینده‌مراقبت از دیابت را دگرگون خواهد کرد اما تغییرات در درمان سریعاً صورت نخواهد گرفت. این مطالعه توسط مرکز دیابت دانشگاه لوند در سوئد و مؤسسه پزشکی مولکولی فنلاند بر روی 160755 بیمار همراه با بررسی دقیق خون آن‌ها انجام شده‌است (Wu et al., 2018). نتایج منتشر شده در دیابت و اندوکریبولوژی لنست نشان می‌دهند که بیماران می‌توانند به پنج گروه متمایز تقسیم شوند (Marinov et al., 2011):

گروه اول: این گروه را دیابت خودایمنی شدید می‌نامند. اینان در سنین پایین مبتلا شده، شاخص توده بدنی آنان (BMI) کم است، کمبود انسولین دارند و دارای اتو آنتی‌بادی‌های ضد گلوتامیک دکربوکسیلاز که شاخصه نوع یک کلاسیک است، می‌باشند لذا به دیابت نوع یک شبیه است و با آن‌که بدن بیمار خصوصیات یک بدن سالم را نشان می‌دهد، ولی به خاطر عارضه بیماری ایمنی توانایی تولید انسولین را ندارد.

گروه دوم: این خوشه دیابت با کمبود شدید انسولین نام دارد. این بیماران به بیماران گروه اول بسیار شبیه هستند ولی با این تفاوت که شاخصه خودایمنی یعنی اتو آنتی‌بادی‌های ضد گلوتامیک دکربوکسیلاز را در سرم خود ندارند؛ لذا آنان توانایی تولید انسولین را ندارد ولی این اختلال از نقص در سیستم ایمنی بدن آن‌ها ناشی نمی‌شود.

هفتمین کنفرانس بین المللی تحقیقات بین رشته‌ای در مهندسی برق، کامپیوتر، مکانیک و مکترونیک در ایران و جهان اسلام



گروه سوم: این خوشه دیابت با مقاومت انسولینی شدید نام گرفته است. بیمارانی که در این گروه قرار می‌گیرند به‌طور کلی دچار عارضه اضافه‌وزن هستند؛ و هرچند بدن این بیماران انسولین کافی تولید می‌کند، اما به دلیل مقاومت انسولینی شدید، بدن دیگر قادر به واکنش‌دهی به این هورمون نیست.

گروه چهارم: این خوشه دیابت خفیف مرتبط با چاقی نام دارد که بیشتر در افرادی با اضافه‌وزن بالا مشاهده می‌شود، ولی این گروه، فاقد مقاومت واضحی نسبت به انسولین هستند و لذا از نظر متابولیک نسبت به بیماران گروه سوم به وضعیت معمولی نزدیک‌تر هستند.

گروه پنجم: دیابت خفیف مرتبط با سن است؛ این خوشه معمولاً در افراد با سن بالا دیده می‌شود و جز سن بالا، مثل خوشه چهارم، اختلال متابولیک واضحی ندارد و لذا علائم این گروه در مقایسه با بیماران سایر گروه‌ها خفیف‌تر است.

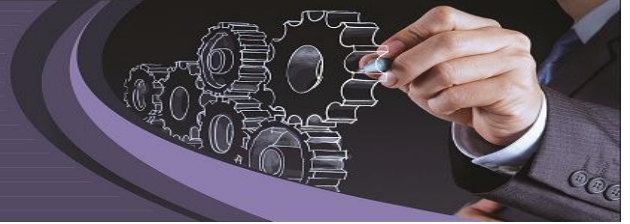
تحول در صنعت سلامت به واسطه این هدف واحد که چگونه سازمان‌های سلامت هزینه‌ها را کاهش و کیفیت را افزایش دهند و همچنان رقابتی باقی بمانند به پیش می‌رود و این مقوله همواره یک چالش بزرگ محسوب می‌شود. بهبود کیفیت در صنعت سلامت را می‌توان به واسطه‌ی نیروهای محرکی که بر آن تاثیرگذار است بهتر تعریف نمود و از جمله این نیروهای محرک داده‌های سلامت است؛ به عبارت دیگر در هر نوع برنامه‌ی بهبود کیفیت متمرکز بر بیمار، داده‌ها قلب آن برنامه به حساب می‌آید. داده‌ها در عصر امروزی یعنی عصر اطلاعات، عمده‌ترین دارایی برای سازمان‌های سلامت بوده و موفقیت سازمان‌های سلامت در گروهی جمع‌آوری، ذخیره و تحلیل آن‌هاست. با این وجود، جمع‌آوری و ذخیره‌ی میزان زیادی از داده‌ها می‌تواند یک نوع اتلاف محسوب شود؛ مگر این که داده‌ها به شکل سودمند استفاده شده و تبدیل به یک منبع مالی برای سازمان گردد. برای تبدیل این ارزش بالقوه به اطلاعات استراتژیک، بسیاری از سازمان‌ها به داده‌های کاوی روی آورده‌اند؛ چرا که به واسطه‌ی داده‌های کاوی امکان کشف روابط، روندها و الگوهای مخفی بین داده‌ها و دستیابی به دانش نوین در زمینه‌ی چالش‌های آشکار و پنهان سازمان میسر خواهد شد.

مفهوم کشف دانش از داده‌ها بیش از یک دهه است که در محیط‌های مالی-تجاری در حال استفاده می‌باشد و در علوم مدیریت ارتباطات، مهندسی، وب‌کاوی، تحلیل جرایم و پزشکی جای خود را باز کرده است. اگرچه کشف دانش باهدف شناسایی اختلاس مالی وارد عرصه‌ی سلامت شد، اما به تدریج در حوزه‌ی بالینی نیز مورد استفاده قرار گرفت. این مهم ناشی از تغییر سریع هوشیاری نسبت به اطلاعات در حوزه‌ی سلامت است. صنعت سلامت به‌طور مستمر در حال تولید میزان زیادی از داده‌ها می‌باشد و افرادی که با این نوع داده‌ها مواجه هستند، دریافته‌اند که بین جمع‌آوری تا تفسیر آنها شکاف وسیعی وجود دارد؛ حوزه‌ی به نسبت جوان و در حال رشد داده‌های کاوی در سلامت از جمله شیوه‌هایی است که می‌تواند این صنعت را از تحلیل عمیق این داده‌ها بهره‌مند سازد و به توسعه‌ی تحقیقات پزشکی و تصمیم‌گیری‌های علمی در زمینه‌ی تشخیص و درمان منتج شود (Marinov et al., 2011).

مبانی داده‌کاوی و هوش مصنوعی در حوضه سلامت

هوش مصنوعی (AI)، هوشی است که توسط ماشین‌ها ظهور پیدا می‌کند، در مقابل هوش طبیعی که توسط جانوران شامل انسان‌ها نمایش می‌یابد. کتاب‌های AI پیشرو، این شاخه را به عنوان شاخه مطالعه بر روی «عوامل هوشمند» تعریف می‌کنند: هر سامانه که محیط خود را درک کرده و کنش‌هایی را انجام می‌دهد که شانسش را در دستیابی به اهدافش بیشینه می‌سازد. برخی از منابع شناخته شده از اصطلاح «هوش مصنوعی» جهت توصیف ماشینی استفاده می‌کنند که عملکردهای «شناختی» را از روی ذهن انسان‌ها تقلید می‌کنند، همچون «یادگیری» و «حل مسئله»، با این حال این تعریف توسط محققان اصلی در زمینه AI رد شده است (Gulati, 2015). اصطلاح هوش مصنوعی برای اولین بار توسط جان مک‌کارتی (پدر علم و دانش تولید ماشین‌های هوشمند) استفاده شد. وی مخترع یکی از زبان‌های برنامه‌نویسی هوش مصنوعی به نام لیسپ (lisp) است. با این عنوان می‌توان به هويت رفتارهای هوشمندانه یک ابزار

هفتمین کنفرانس بین المللی تحقیقات بین رشته‌ای در مهندسی برق، کامپیوتر، مکانیک و مکترونیک در ایران و جهان اسلام



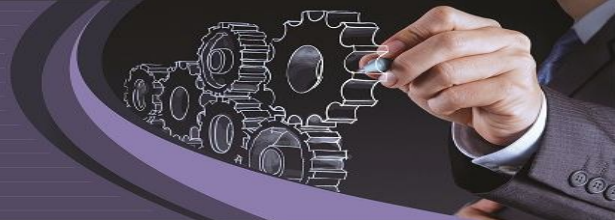
مصنوعی پی برد (ساخته دست بشر، غیرطبیعی، مصنوعی) حال آنکه هوش مصنوعی به عنوان یک اصطلاح عمومی پذیرفته شده که شامل محاسبات هوشمندانه و ترکیبی (مربک از مواد مصنوعی) است. داده کاوی بخشی از هوش مصنوعی است که به مفهوم استخراج اطلاعات نهان یا الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ گفته می‌شود. بسیاری از مردم داده کاوی را مترادف واژه‌های رایج کشف دانش در پایگاه داده‌ها می‌دانند. داده‌کاوی، پایگاه‌ها و مجموعه حجیم داده‌ها را در پی کشف و استخراج، مورد تحلیل قرار می‌دهد. این‌گونه مطالعات و کاوش‌ها را به واقع می‌توان همان امتداد و استمرار دانش کهن و همه‌جا گیر آمار دانست. تفاوت عمده در مقیاس، وسعت و گوناگونی زمینه‌ها و کاربردها، و نیز ابعاد و اندازه‌های داده‌های امروزی است که شیوه‌های ماشینی مربوط به یادگیری، مدل‌سازی، و آموزش را طلب می‌نماید (Bhavsar and Waghmare, 2013). داده کاوی به بهره‌گیری از ابزارهای تجزیه و تحلیل داده‌ها به منظور کشف الگوها و روابط معتبری که تاکنون ناشناخته بوده‌اند اطلاق می‌شود. این ابزارها ممکن است مدل‌های آماری، الگوریتم‌های ریاضی و روش‌های یاد گیرنده باشند که کار این خود را به صورت خودکار و بر اساس تجربه‌ای که از طریق شبکه‌های عصبی یا درخت‌های تصمیم‌گیری به دست می‌آورند بهبود می‌بخشد. داده کاوی منحصر به گردآوری و مدیریت داده‌ها نبوده و تجزیه و تحلیل اطلاعات و پیش‌بینی را نیز شامل می‌شود برنامه‌های کاربردی که با بررسی فایل‌های متن یا چند رسانه‌ای به کاوش داده‌ها می‌پردازند پارامترهای گوناگونی را در نظر می‌گیرد که عبارت اند از (Vijiyarani and Sudha, 2013):

قواعد انجمنی (Association): الگوهایی که بر اساس آن یک رویداد به دیگری مربوط می‌شود مثلاً خرید قلم به خرید کاغذ.
ترتیب (Sequence): الگویی که به تجزیه و تحلیل توالی رویدادها پرداخته و مشخص می‌کند کدام رویداد، رویدادهای دیگری را در پی دارد مثلاً تولد یک نوزاد و خرید پوشک.

پیش‌بینی (Prediction): در پیش‌بینی هدف پیش‌بینی یک متغیر پیوسته می‌باشد. مانند پیش‌بینی نرخ ارز یا هزینه‌های درمانی.
رده‌بندی یا طبقه‌بندی (Classification): فرایندی برای پیدا کردن مدلی است که رده‌های موجود در داده‌ها را تعریف می‌نماید و متمایز می‌کند، با این هدف که بتوان از این مدل برای پیش‌بینی رده رکوردهایی که برچسب رده آن‌ها (متغیر هدف) ناشناخته می‌باشد، استفاده نمود. در حقیقت در رده‌بندی بر خلاف پیش‌بینی، هدف پیش‌بینی مقدار یک متغیر گسسته است. روش‌های مورد استفاده در پیش‌بینی و رده‌بندی عموماً یکسان هستند.

خوشه‌بندی (Clustering): گروه‌بندی مجموعه‌ای از اعضاء، رکوردها یا اشیاء به نحوی که اعضای موجود در یک خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را به اعضای خوشه‌های دیگر داشته باشند.

مصورسازی (Visualization): مصورسازی داده‌ها یکی از قدرتمندترین و جذابترین روش‌های اکتشاف در داده‌ها می‌باشد. بطورکلی هدف داده کاوی، یادگیری و آموختن از داده‌ها است و بر این اساس و 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 بیمه‌ی یک موسسه‌ی 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 شیوه‌ی یادگیری فاقد نظارت، ویژگی‌ها و مدل‌های خطاهای مطالبات شناخته شده نیست، اما الگوها و خوشه‌های حاصل از داده کاوی منجر به کشفیات جدید می‌شود (Das et al., 2018). با استفاده از داده کاوی و مدل‌سازی داده‌ها مبتنی بر هوش مصنوعی می‌توان بیماران با شرایط پرخطر را شناسایی کرد. در واقع داده‌کاوی با ارائه‌ی اطلاعات به ارایه دهندگان مراقبت، آن‌ها را در شناسایی بیماران پرخطر به گونه‌ای که بتوان کیفیت مراقبت آن‌ها را بهبود داد و از مشکلات آتی آنها جلوگیری کرد، کمک می‌کند و باطراحی مداخله‌ی مناسب منجر به کاهش پذیرش‌های بیمارستانی می‌شود. به عنوان مثال تکنیک‌های مدل‌سازی پیش‌بینی



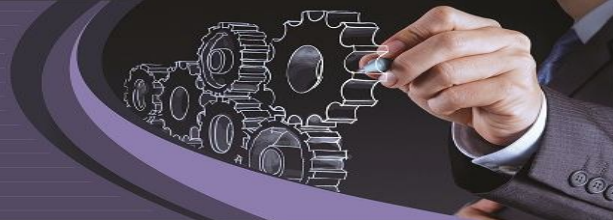
کننده ی داده کاوی در رابطه با مدیریت بیماری دیابت منجر به ارتقای کیفیت و کاهش هزینه ی بیماران مبتلا به دیابت می‌شود (Jacob and Ramani, 2012).

تشخیص دیابت با داده کاوی

حوزه بهداشت و درمان صورت گرفته است و با نام داده کاوی بالینی شناخته می‌شود که با توسعه و پیشرفت ابزارهای فن آوری اطلاعات و ارتباطات می‌تواند راهکار نوین و مفیدی برای کشف الگوهای موجود در داده های سلامت باشد. سیستم های مراقبت های بهداشتی مملو از اطلاعاتی است که دانش قابل توجهی در آن نهفته است بنابراین به تکنیک ها و ابزارهایی برای استخراج اطلاعات مفید از این مجموعه داده های بزرگ نیاز است. اولین مزیت استفاده از این روش، امکان تحقیق برای کشف الگوهای پنهان در مجموعه داده های موجود در حوزه بهداشت و درمان می باشد که از این الگوها می توان برای تشخیص بیماری ها و همچنین درمان آن ها استفاده نمود (Thakkar et al., 2021). در این میان بیماری دیابت به چند دلیل یکی از بیماری های مناسب برای کاربرد داده کاوی در پزشکی می باشد. دیابت از جمله بیماری هایی است که عوارض جانبی متعدد و دردناکی همچون نارسایی قلب و کلیه، نابینایی و حتی قطع عضو را در پی دارد که هر کدام از آن ها به تنهایی هزینه های مادی و معنوی زیادی را برای فرد و جامعه همراه دارد. از طرفی به دلیل شیوع بالای این بیماری، شناسایی و کشف الگوهای بیماری و روش درمان آن می تواند موجب صرفه جویی قابل ملاحظه ای در زمان و هزینه های اقتصادی برای مؤسسات ارائه کننده سلامت و همچنین بیماران شود. از طرفی به دلیل تاثیر عوامل متعدد در این بیماری، متغیرهای بسیاری در تصمیم گیری های پزشکی مربوط به آن در نظر گرفته می شوند. بنابراین پیاده سازی روشی که بتواند امکان تشخیص صحیح ابتلا به دیابت و یا عدم ابتلا را مشخص کند می تواند گام مهمی در پیشگیری و کنترل این بیماری به خصوص در مراحل ابتدایی آن باشد. که این موارد به همراه وجود بانک های اطلاعاتی متعددی که اطلاعات مربوط به بیماران مبتلا به دیابت را در دوره های زمانی متفاوت ثبت و نگهداری می کنند. مانند مجموعه داده های PIMA (داده های مربوط به سرخپوستان آمریکای شمالی)، موجب می شود بیماری دیابت به عنوان یکی از حوزه های مناسب برای به کارگیری داده کاوی باشد (Sa'di et al., 2015).

در واقع داده کاوی می تواند به عنوان 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. بر همین اساس، مطالعات مختلفی در سال های اخیر با هدف بررسی و به کارگیری الگوریتم های پرکاربرد داده کاوی در بیماری دیابت صورت گرفته که در ادامه با معرفی هر یک از این روش ها و توضیح مختصری در مورد آن ها، نتایج مربوط به نمونه هایی از مطالعاتی که به صورت کاربردی از این روش ها استفاده کرده اند، ارائه می شود (Islam et al., 2020). مشکل عمدهای که در رابطه با بیماری دیابت وجود دارد عدم تشخیص به موقع و یا به طور کلی ضعف در تشخیص این بیماری است که این ضعف نیز به دلیل عدم انتخاب الگوی مناسب توسط پزشک و یا عدم استفاده مناسب از الگوهای استاندارد است. بنابراین پیاده سازی روشی که بتواند هر فرد را در تشخیص صحیح ابتلا یا عدم ابتلا به این بیماری یاری رساند می تواند گام مهمی در جهت پیشگیری و کنترل این بیماری به خصوص در مراحل ابتدایی آن باشد (Kim et al., 2012). واسان و همکاران (۲۰۰۶) توانستند بر پایه روش های شبکه عصبی مصنوعی، درخت تصمیم، رگرسیون و قواعد وابستگی بر پایه عکس های سه بعدی و دو بعدی بدن با دقت ۸۹ درصد پیش بینی کنند که آیا فرد مورد نظر به بیماری دیابت نوع دوم مبتلا است یا خیر. همچنین صنعتی (۲۰۰۹) با استفاده از روش ماشین بردار پشتیبان توانستند با دقت ۹۳ درصد بیماری دیابت نوع دوم را در افراد تشخیص دهند. داده های آن ها مربوط به ۷۶۸ بیمار می شد و از هشت متغیر از جمله فشار خون افراد و میزان انسولین تزریقی برای پیش بینی خود استفاده نمودند. براکات و همکاران (۲۰۱۰) با استفاده از روش ماشین

هفتمین کنفرانس بین المللی تحقیقات بین رشته‌ای در مهندسی برق، کامپیوتر، مکانیک و مکترونیک در ایران و جهان اسلام



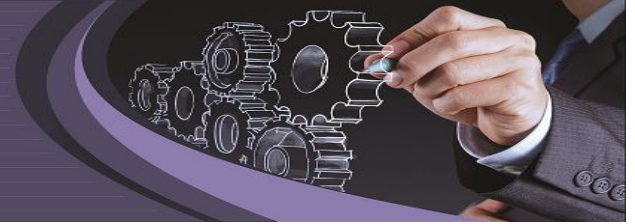
بردار پشتیبان توانستند دقت تشخیص بیماری را بهبود دهند و با دقت ۹۴ درصد بیماری دیابت نوع دوم را در افراد تشخیص دهند. آن‌ها داده‌های تحقیق خود را از داده‌های مربوط به ۶۸۲ نفر مراجعه کننده استخراج کردند و متغیرهای آن‌ها شامل: جنسیت، شاخص نسبت وزن به قد (BMI) فشارخون، کلسترول و قندخون بوده است. در سال ۲۰۱۰ عده‌ای از محققان تایلندی با استفاده از درخت تصمیم توانستند با دقت بیش از ۹۰ درصد سندرم متابولیک را در افراد تشخیص دهند. آن‌ها از داده‌های مربوط به ۵۶۳۸ نفر استفاده کرد (Komi et al., 2017). لیو و همکاران (۲۰۱۴)، 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 خودمراقبتی برای بیماران مبتلا به دیابت نوع دوم باشد و همچنین این سیستم با یکپارچه‌سازی داده‌کاوی، محاسبات همراه و دانش پزشکی در یک سیستم هوشمند موبایل، می‌تواند برای افراد مبتلا به بیماری‌های مزمن مانند دیابت مفید باشد. حبیبی و همکاران (۲۰۱۵) 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 ۲۰۰۹ تا ۲۰۱۱ 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 ۷۱٪ بوده است.

بحث و نتیجه‌گیری

یکی از مهمترین موضوعات چالش برانگیز در مراقبت سلامت، تغییر شکل داده‌های بالینی خام به اطلاعات معنی‌دار به دنبال تولید مستمر انبوهی از داده‌ها است چرا که بسیاری از سازمان‌های مراقبت سلامت با وجود غنای داده با فقر دانش روبرو شده‌اند. به عبارت دیگر رویارویی با مجموعه‌های عظیم داده‌ها و توسعه‌ی پایگاه‌های داده‌ها نسبت به دهه‌های گذشته نیازهای جدیدی را مانند خلاصه‌سازی خودکار داده‌ها، استخراج اطلاعات ذخیره شده و کشف الگوها از داده‌های خام به وجود آورده است که داده‌کاوی نمونه‌ای از آن می‌باشد. استخراج اطلاعات و دانش از داده‌ها مفهومی دیرینه در مطالعات علمی و پزشکی می‌باشد و آنچه که جدید است هم‌گرایی و اشتراک چندین رشته و فن‌آوری‌های متناظر آنها است که فرصت منحصر به فردی را برای داده‌کاوی ایجاد کرده است. داده‌کاوی با ایجاد پزشکی مبتنی بر شواهد نقش حیاتی در سلامت داردرومنجر به کشف دانش جدید، سودمند و ماندگار در پایگاه‌های داده‌ای سازمان‌های سلامت می‌شود؛ چرا که برای دستیابی به پزشکی مبتنی بر شواهد باید از شناسایی شکاف و خلاء دانش در فرایندهای مراقبت سلامت کنونی شروع کرد و سپس به دنبال بهترین ادله بود، در قدم بعدی باید به بررسی صحیح بودن و معتبر بودن اقدامات شناسایی شده در بهترین ادله پرداخت و در قدم آخر این ادله را بر روی بیماران اجرا کرد. داده‌کاوی دستیابی به اولین گام در این زمینه را هموار می‌سازد.

منابع

- Abdollahi, J., Nouri-Moghaddam, B. (2022). Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. *Iran Journal of Computer Science*, Vol. 52, pp. 1-16.
- Barakat, N., Bradley, A., Barakat, M. (2010). Intelligent Support Vector Machines for Diagnosis of Diabetes Mellitus. *IEEE Transactions on information technology in biomedicine*, Vol. 14, No. 4, pp. 1114-1120.



- Bhavsar, Y.B., Waghmare, K.C. (2013). Intrusion detection system using data mining technique: Support vector machine. *International Journal of Emerging Technology and Advanced Engineering*, Vol. 3, No. 3, pp. 581-586.
- Das, H., Naik, B., Behera, H.S. (2018). Classification of diabetes mellitus disease (DMD): a data mining (DM) approach. In *Progress in computing, analytics and networking*, Springer, pp. 539-549.
- Gulati, H. (2015). Predictive analytics using data mining technique. In: *IEEE 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 713-716.
- Habibi, S., Ahmadi, M., Alizadeh, S. (2015). Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining. *Global Journal of Health Science*, Vol. 7, No. 5, pp. 304-10.
- Islam, M. M., Ferdousi, R., Rahman, S., Bushra, H.Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *Computer Vision and Machine Intelligence in Medical Image Analysis*, Vol. 7, pp. 113-125.
- Jacob, G., Ramani, R. (2012). Data Mining in Clinical Data Sets: A Review. *International Journal of Applied Information Systems (IJ AIS)*, Vol. 4, No. 6, pp. 15-26.
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, Vol. 15, pp. 104-116.
- Kim, H.S., Shin, A.M., Kim, M.K., Kim, Y.N. (2012). Comorbidity study on type 2 diabetes mellitus using data mining. *The Korean journal of internal medicine*, Vol. 27, No. 2, pp. 197.
- Komi, M., Li, J., Zhai, Y., Zhang, X. (2017). Application of data mining methods in diabetes prediction. In: *IEEE 2017 2nd international conference on image, vision and computing (ICI VC)*, pp. 1006-1010.
- Luo, Y., Ling, C., Schuurman, J., Petrella, R. (2014). GlucoGuide: An Intelligent Type-2 Diabetes Solution Using Data Mining and Mobile Computing. *IEEE International Conference on Data Mining Workshop*, Shenzhen: IEEE.
- Marinov, M., Mosa, A. S. M., Yoo, I., Boren, S. A. (2011). Data-mining technologies for diabetes: a systematic review. *Journal of diabetes science and technology*, Vol. 5, No. 6, pp. 1549-1556.
- Nandita, R., Rao, M. (2013). Association Rule Mining on Type 2 Diabetes using FP-growth association rule. *International Journal of Engineering And Computer Science*, Vol.2, No. 8, pp. 2481-85.
- Sa'di, S., Maleki, A., Hashemi, R., Panbechi, Z., Chalabi, K. (2015). Comparison of data mining algorithms in the diagnosis of type II diabetes. *International Journal on Computational Science & Applications (IJCSA)*, Vol. 5, No. 5, pp. 1-12.
- Santi, WP. (2009). A new smooth support vector machine and its applications in diabetes disease diagnosis. *J Comput Sci*, Vol. 5, 1003-1008.
- Sheikhpour, R., AghaSarram, M. (2014). Diagnosis of Diabetes Using an Intelligent Approach Based on Bi-Level Dimensionality Reduction and Classification Algorithms. *Iranian Journal of Diabetesband Onesity*, Vol. 6, No. 2, pp. 74-84.
- Thakkar, H., Shah, V., Yagnik, H., Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical eHealth*, Vol. 4, pp. 12-23.
- Vijayarani, S., Sudha, S. (2013). Disease prediction in data mining technique—a survey. *International Journal of Computer Applications & Information Technology*, Vol. 2, No. 1, pp. 17-21.
- Wasan, S., Bhatnagar, V., Kaur, H. (2006). The impact of data mining techniques on medical diagnostics *Data Science Journal*, Vol. 5, pp. 119-26.
- Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, Vol. 10, pp. 100-107.