مطالعات علوم زیستی و زیست فناوری
دوره ۸، شماره ۴، زمستان ۱۴۰۱
صفحات ۳۴- ۲۵

Online ISSN: ٤٤٩٣-٢٤٧٦
Print ISSN: ٢٠٢٠-٢٥٣٨
www.irijournals.com

بررسی بیماری‌های مزمن کلیوی بر اساس تکنیک‌های داده‌کاوی: یک مطالعه مروری

نوید کیانوش مقدم[۱]، مهسا اکبری عریانی[۲]، فرزانه فرزام[۳]

[۱] گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه فنی و حرفه‌ای، تهران، ایران

[۲]استادیار گروه آسیب‌شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

[۳]دانشکده پزشکی، دانشگاه علوم پزشکی نیشابور، نیشابور، ایران

چکیده

بیماری نارسایی کلیه یک بیماری رایج در سراسر جهان است. در این بیماری عملکرد کلیه به‌تدریج کاهش می‌یابد و اگر بدتر شود منجر به نارسایی حاد کلیوی شود. در این مطالعه برای پیش‌بینی بیماری نارسایی کلیه دیدگاه‌های مختلف موردبررسی قرار می‌گیرند، همچنین به تحلیل مجموعه داده‌های استفاده‌شده در پیش‌بینی این بیماری بر اساس روش‌های داده‌کاوی می‌پردازیم. این اختلال حدود ۶۰ درصد از کل مرگ‌ومیرها را در سراسر جهان را تشکیل می‌دهند. ۸۰ درصد مرگ‌ومیر ناشی از بیماری‌های مزمن کلیوی نیز در سراسر جهان در کشورهای کم‌درآمد و متوسط رخ می‌دهد. داده‌کاوی فرآیند استخراج اطلاعات از مجموعه داده‌های فراوان است. تکنیک‌های عمده داده‌کاوی مانند خوشه‌بندی، طبقه‌بندی، تجزیه‌وتحلیل ارتباط، رگرسیون، سری زمانی و تجزیه‌وتحلیل توالی برای پیش‌بینی بیماری‌های کلیوی در برخی مطالعات استفاده‌شده است. تکنیک‌هایی که تاکنون معرفی شده‌اند دارای اشکالات جزئی در کیفیت پیش‌پردازش یا در هر مرحله دیگری بودند. در این مقاله، تکنیک‌های مختلف داده‌کاوی برای پیش‌بینی بیماری‌های مزمن کلیوی بررسی شده و مشکلات عمده به‌اختصار توضیح داده شده است.

کلمات کلیدی: بیماری مزمن کلیوی، تکنیک داده‌کاوی، مطالعه مروری

مقدمه

پیشرفت سریع به دلیل حجم بالای داده‌های بیولوژیکی موجود، یک زمینه به‌منظور طبقه‌بندی، مدیریت مؤثر و انتقال داده‌های خام به اطلاعات معنادار را ایجاد کرده است (۱). ظهور این حجم عظیم از اطلاعات و داده‌ها نیازمند الگویی از محاسبات مدرن است تا بتواند این اطلاعات را به‌صورت داده‌های معنادار مورد ارزیابی قرار دهد (۲). یکی از مواردی که در این زمینه بیان شده است، این است که می توان با استفاده از الگوریتم های داده‌کاوی به ارزیابی این دسته از داده‌ها پرداخت (۳, ۴).

یادگیری روش‌های تحلیل داده بخش عمده ای از روش‌های داده‌کاوی می باشد، و اهمیت آخرین پیشرفت ها در زمینه بیوانفورماتیک، زیست شناسی محاسباتی و کاربرد روش‌های یادگیری ماشین در مسائل برجسته در زیست شناسی و رفتار انسان را نشان می‌دهد (۵). الگوریتم‌ها و تکنیک‌های ریاضی به ما این امکان را می‌دهند که فراتر از تصویرسازی صرفا با استفاده از داده‌ها و نتایج منطقی در قالب مدل‌های قابل آزمایش ریاضی، نتایج قابل ارجاعی را ایجاد کنیم (۶, ۷).

از آنجایی که زندگی روزمره مردم مدرن تر می شود و طول عمر آنها در جامعه افزایش می‌یابد، بیماری مزمن کلیه (CKD) نیز شایع شده و منجر به تخریب عملکرد کلیه شده است. ابتلا به CKD ممکن است باعث کاهش توانایی فعالیت و افت کیفیت زندگی گردد (۸). همچنین این اختلال به سرعت منجر به سایر بیماری‌های مزمن مانند فشار خون بالا، کم خونی، استخوان های ضعیف به دلیل سلامت تغذیه ضعیف و آسیب عصبی می شود. در این میان، بیماری کلیوی خطر ابتلا به بیماری‌های قلبی و عروقی را به حداکثر می‌رساند (۱, ۹).

بیماری مزمن کلیه حتی باعث بیماری‌های مزمن دیگری مانند دیابت، فشار خون بالا و سایر اختلالات می شود. گروه های پرخطر به عنوان افراد مبتلا به دیابت، فشار خون بالا و ارثی طبقه‌بندی می شوند (۱۰). رهایی از بیماری مزمن کلیوی از طریق تشخیص زودهنگام و درمان مناسب، پس از مشاهده پیشرفت بیماری ممکن است تا حد زیادی منجر به نارسایی کلیه شود. مجموعه داده‌های مرتبط با CKD از مجموعه UCI گرفته شده است (۱۱). با توجه به شیوع بالای بیماری‌های کلیوی، و وجود اطلاعات فراوان در این زمینه، اهمیت دسته بندی این اطلاعات به‌منظور ارزیابی دقیقتر بیماران و تشخیص زودهنگام آن ها مشخص می شود (۵, ۱۲). بر این اساس در مطالعه اخیر به ارزیابی عوامل حمایت کننده از بیماری‌های مزمن کلیوی و اهمیت داده‌کاوی در این بیماری پرداخته ایم.

بررسی مطالعات

برای اندازه‌گیری شدت های مختلف CKD از عوامل تأثیرگذار به عنوان ورودی داده برای مدل‌سازی استفاده می‌شود. در مرحله بعد، تشخیص و طبقه‌بندی CKD و ارزیابی خطر انجام می شود. سه مدل مختلف برای تشخیص استفاده می‌شوند. بیماری کلیوی بر آسیب کلیه تأثیر می گذارد و خون را فیلتر نمی کند. این آسیب می تواند باعث باقی ماندن مواد زائد در بدن شود (۱۰, ۱۳). در اکثر بیماران، آسیب کلیوی به آرامی در طول سال ها رخ می‌دهد، که به دلیل دیابت یا فشار خون بالا به نام بیماری مزمن کلیوی است. هنگامی که فردی به دلیل بیماری، آسیب دیدگی، مصرف دارو خاص، تغییر ناگهانی در عملکرد کلیه داشته باشد، آسیب حاد کلیه (AKI) نامیده می شود (۳). این می تواند در فردی با کلیه های طبیعی یا در هر فردی که قبلاً از مشکلات کلیوی رنج می برد رخ دهد. به نظر می رسد یک مشکل رو به رشد است. ۱۰ درصد از جمعیت جهان به این بیماری مبتلا هستند (۵, ۱۴). سالانه

میلیون ها نفر به دلیل بیماری مزمن کلیه جان خود را از دست می‌دهند زیرا تحت درمان مناسب قرار نمی‌گیرند. مطالعه بار جهانی بیماری که در سال ۲۰۱۰ انجام شد، مشخص کرد که بیماری مزمن کلیوی در حدود سال ۱۹۹۰ در رتبه ۲۷ در فهرست تعداد کل مرگ‌ومیر در سراسر جهان قرار دارد و می تواند در سال ۲۰۱۰ به رتبه هجدهم برسد (۱۵, ۱۶). عوامل خطر بیماری کلیوی، شامل سیگار کشیدن، چاقی، کلسترول بالا، دیابت (نوع I و II)، بیماری خودایمنی، بیماری انسدادی کلیه، انسداد مثانه ناشی از هیپرپلازی خوش خیم پروستات، تصلب شرایین، سیروز و نارسایی کبد، تنگی شریان، سرطان کلیه، سرطان مثانه، سنگ کلیه و عفونت های مرتبط هستند (۱۷). بیماری‌های کلیوی با استفاده از الگوریتم SVM و شبکه ارتباطی بر اساس دقت و زمان اجرای پیش‌بینی اختلال کلیوی و مقایسه آن ها انجام می شود (۷, ۱۴).

نتایج نشان می‌دهد که ANN با کاهش زمان اجرا بهتر عمل می‌کند. مجموعه داده‌های CKD از مخزن UCI گرفته شده است و عملکرد با استفاده از الگوریتم هایی مانند Naive Bayes، Multilayer Perceptron، SVM، J48، Conjunctive و Decision tree ارزیابی می شود (۱۸). این نشان می‌دهد که الگوریتم پرسپترون دقت بالاتر طبقه‌بندی بیماری‌های کلیوی را نشان می‌دهد. الگوریتم خوشه‌بندی K-Means نیز برای طبقه‌بندی خوشه‌هایی با احتمال متغیر ابتلا به CKD فرموله شده است (۱۹). همچنین الگوریتم‌های یادگیری ماشینی مانند AD Trees، J48، KStar، Naive Bayes، الگوریتم‌های Random Forest برای پیش‌بینی بیماری کلیوی استفاده می‌شوند (۶, ۱۵). از سوی دیگر بیان شده است که نتیجه عملکرد Naive Bayes میزان دقت بهتری را در مقایسه با سایر الگوریتم‌ها نشان می‌دهد (۴, ۲۰).

مجموعه داده‌های موجود در زمینه بیماری مزمن کلیه برای پیش‌بینی وضعیت بیماران مبتلا به نارسایی مزمن کلیه و افراد عادی استفاده می شود. سه مدل مختلف شبکه داده برای پیش‌بینی بیماری مزمن کلیوی اجرا شده‌اند که شامل شبکه های پروپاگیشن، شبکه جنرالایزد و شبکه عصبی مدولار است. هر سه مدل دقت بیش از ۸۵٪ ارائه می‌دهند و در مقایسه با سایر مدل ها، شبکه پروپاگیشن بالاترین دقت را دارد (۲). بیماری مزمن کلیوی با استفاده از شش الگوریتم داده‌کاوی مختلف شامل طبقه‌بندی‌کننده‌های تصادفی، بهینه‌سازی حداقل متوالی، نایو بیز، تابع پایه شعاعی، طبقه‌بندی کننده پرسپترون چندلایه و لجستیک ساده پیش‌بینی می‌شود (۳, ۴).

در مجموع ۴۰۰ رکورد برای آموزش در زمینه بیماری ها بیان شده است، که در این میان، Random Forest عملکرد بهتری دارد. زمان واقعی جمع آوری شده از مجموعه داده‌ها شامل بسیاری از داده‌های نامتعادل است. برای پرداختن به این موضوع یک الگوریتم متعادل کننده مجدد به نام SMOTE برای پیش‌بینی بیماری کلیوی بیماران ارائه شده است (۱۵). برای بهبود دقت نتیجه پیش‌بینی، الگوریتم‌های طبقه‌بندی مانند Naive Bayes و الگوریتم‌های Decision tree مقایسه می‌شوند؛ اما به نظر می رسد دقت طبقه‌بندی Decision tree ۹۱ درصد باشد (۱۳, ۲۱).

عوامل خطر CKD

افزایش اسیب پذیری

برخی عوامل منجر به افزایش آسیب پذیری کلیه نسبت به عوامل اسیب رسان می شود. این علائم در افراد مسن دیده می شود، و برخی عوامل ارثی نیز در آن موثرند (۲۲). کاهش توده کلیه، وزن پایین کودک، درآمد کم و سطح پایین تحصیلات در آن موثر شناخته شده است (۲۳, ۲۴).

عوامل محرک

عواملی که مستقیماً در آسیب کلیه دخیل هستند را مشخص می کند. این عوامل شامل دیابت قندی، فشار خون بالا، بیماری‌های خودایمنی، عفونت‌های سیستمیک، عفونت‌های مجاری ادراری، سنگ‌های ادراری، عوامل انسدادی دستگاه ادراری تحتانی و داروهای سمی است (۲۵, ۲۶).

عوامل پیشرفت بیماری کلیوی

به محض شروع آسیب کلیوی، برخی عوامل منجر به تشدید آسیب کلیه و کاهش سریع عملکرد کلیه می شود؛ و علایم مرتبط با آن شامل پروتئینوری سطح بالا، فشار خون بالا، کنترل قند خون کمتر به دلیل دیابت و سیگار کشیدن است (۲۷).

عوامل ESRD

افزایش موربیدیتی و مرگ‌ومیر به دلیل نارسایی کلیه، کاهش دوز دیالیز، دسترسی موقت عروقی، کم خونی، سطح آلبومین پایین سرم و ارجاع دیرهنگام جهت دیالیز رخ می‌دهد (۱۶, ۲۸).

کارایی داده‌کاوی

داده‌کاوی را می توان به عنوان جستجو به‌منظور ورود داده‌های آماری به سیستم تعریف کرد. جستجو داده‌کاوی شامل وظایف اولیه داده‌کاوی است (۲۹). در این فرایند به کاربر اجازه داده می شود که به‌صورت تعاملی با سیستم داده‌کاوی ارتباط برقرار کند و فرآیند استخراج داده‌ها را انجام دهد، یا یافته ها را از زوایای مختلف بررسی کند. داده‌های مورد نیاز را می توان با استفاده از پایگاه داده یا مجموعه داده‌هایی که کاربر می خواهد استخراج کند، ارزیابی کرد (۳۰).

داده‌کاوی شامل ویژگی های پایگاه داده یا ابعاد داده موردبررسی است. دانش را می توان با استفاده از توصیف، تبعیض، تجزیه‌وتحلیل ارتباط یا همبستگی، طبقه‌بندی، پیش‌بینی، خوشه‌بندی، تجزیه‌وتحلیل استخراج کرد. دانش پس زمینه در فرآیند کشف برای ارزیابی الگوهای یافت شده استفاده می شود. سلسله مراتب مفهومی شکل محبوبی از دانش پس‌زمینه است که به داده‌های استخراج شده در سطوح مختلف اجازه ارزیابی تحلیلی می‌دهد. انواع مختلف دانش ممکن است معیارهای جذابیت متفاوتی مانند حمایت و اطمینان از قوانین (Charu) داشته باشند (۳۱-۳۳).

مزایا و معایب داده‌کاوی در حوزه پزشکی

مزایا

- ✓ داده‌کاوی تقلب و سوء استفاده را در مراقبت های بهداشتی شناسایی می کند.
- ✓ به پزشکان کمک می کند تا درمان‌های مؤثر و بهترین شیوه‌ها را استفاده کنند.
- ✓ باعث می شود بیماران از خدمات مراقبت های بهداشتی بهتر و مقرون به صرفه تر استفاده کنند (۱۶).
- ✓ افزایش سرعت کار با مجموعه داده‌های بزرگ و تولید سریع گزارش، تجزیه‌وتحلیل سریع تر داده‌ها بهبود کارایی و کاهش هزینه (۲۱).
- ✓ داده‌کاوی می تواند دانش پیش‌بینی را از پایگاه های داده بزرگ استخراج کند (۶, ۸, ۲۳).

معایب

- ✓ تفسیر ناهمگون با طبقه‌بندی ریاضی ضعیف.
- ✓ مسائل اخلاقی، حقوقی و اجتماعی.
- ✓ مسائل مربوط به مالکیت داده‌ها.
- ✓ حریم خصوصی و امنیت مربوط به مدیریت داده‌های انسانی (۵, ۳۴).
- ✓ مسائل مربوط به حریم خصوصی و مسائل امنیتی و
- ✓ سوء استفاده از داده‌ها یا ارائه اطلاعات نادرست است (۳۴-۳۶).

الگوریتم ها و تکنیک‌های داده‌کاوی

الگوریتم داده‌کاوی یک روش کاملاً تعریف شده است که داده‌ها را به عنوان ورودی می گیرد و به عنوان خروجی اطلاعات تحلیل شده تولید می کند. این الگوریتم ها شامل الگوهایی در قالب مدل هایی ارائه شده است. در این زمینه چندین الگوریتم و تکنیک مانند طبقه‌بندی، خوشه‌بندی، پیش‌بینی، قوانین انجمن، شبکه ها و ... برای کشف دانش و تحلیل اطلاعات از پایگاه های داده وجود دارد (۴, ۳۴).

طبقه‌بندی

طبقه‌بندی ساده ترین و یکی از روش‌های محبوب داده‌کاوی است. جایی که اشیاء تقسیم می شوند و به گروه های دیگر به نام کلاس ها اختصاص می یابند (۹). هر شی باید دقیقاً به یک کلاس اختتصاص داده شود و هیچ داده ای نباید به بیشتر از یک کلاس اختصاص داده شود و یا در هیچ کلاسی تقسیم بندی نشود (۱۱, ۳۳). الگوریتم های طبقه‌بندی عبارتند از:

الگوریتم Naive Bayes

الگوریتم Naive Bayes احتمالات قبلی و شرطی را در یک فرمول ترکیب می کند، به طوری که می توان از آن برای محاسبه احتمال هر یک از طبقه‌بندی های ممکن استفاده کرد (۶).

الگوریتم Nearest Neighbor

این نوع الگوریتم مبتنی بر فاصله است که عمدتاً زمانی استفاده می شود که همه مقادیر پیوسته هستند، و داده‌ها ممکن است با توجه به ویژگی های طبقه‌بندی اصلاح شوند. بر این اساس داده‌هایی که به یک طبقه نزدیک هستند در ان قسمت طبقه‌بندی می شوند (۲۹).

الگوریتم مبتنی بر Decision Tree

این رویکرد برای طبقه‌بندی و مدل‌سازی داده‌هایی که امکان تقسیم بندی به‌صورت یک شاخه هایی از یک ستون اصلی وجود دارد مفید است (۱۲, ۳۳).

خوشه‌بندی

این داده‌ها به‌منظور ارزیابی گروه هایی از داده‌ها به عنوان روش‌های مبتنی بر پارتیشن، سلسله مراتبی و تجمعی استفاده می شود.

روش مبتنی بر پارتیشن

این روش گروه بندی کلی پایگاه داده، ورودی داده‌ها را در پارتیشن های k به عنوان خوشه بررسی می کند. خوشه ها با بهینه سازی یک پارامتر برای محاسبه فاصله بین خوشه ها بر اساس اندازه گیری میزان تطابق بین آن ها ایجاد می شوند. این روش متغیر های مشابهی را که در یک خوشه قرار می‌گیرند را بررسی می‌کند و اشیاء نامرتبط را در زیر خوشه‌های مختلف گروه‌بندی می‌کند. این روش شامل خوشه‌بندی K-Means و K-Medoids است (۲۰, ۲۹).

الگوریتم خوشه‌بندی سلسله مراتبی

تکنیک خوشه‌بندی سلسله مراتبی بر اصل تجزیه پایگاه‌های داده به‌صورت بالا به پایین یا پایین به بالا متمرکز است که این تکنیک به انواع تقسیم کننده و انباشته تقسیم می شود (۱۶).

پیش‌بینی

این تکنیک مقادیر داده‌های وارد نشده و میس شده را پیش‌بینی کرده یا حدس می‌زند. رگرسیون خطی ساده ترین نوع در میان انواع مختلف رگرسیون است که در آن داده‌ها با خطوط مستقیم مدل می شوند. داده‌هایی که با خطوط مستقیم مدل‌سازی می‌شوند را می‌توان به عنوان رشد داده‌ها به شکل خطی یا دنبال کردن یک خط مستقیم تفسیر کرد. ضرایب رگرسیون با استفاده از روش حداقل مربعات برآورد می شوند (۷, ۲۹).

بحث

داده‌کاوی در استخراج اطلاعات از مجموعه وسیع پایگاه داده‌ها اهمیت بسیار زیادی دارد. اگرچه ما به تکنیکی برای استفاده هوشمندانه از روش‌های داده‌کاوی به‌منظور استخراج اطلاعات دقیق از داده‌های فراوان نیاز داریم، و استفاده از این روش ها ساده نبوده و در مطالعات بیان شده مشکلات موجود در این زمینه بیان شده و به برخی روش‌های کاربردی در این زمینه اشاره شده است (۳۷). برای استفاده از تکنیک‌های داده‌کاوی، درک ماهیت داده‌ها بسیار مهم است. مهمترین مرحله بیان شده شامل پیش‌پردازش داده‌ها است، و اگر پیش‌پردازش داده‌ها به درستی صورت نگیرد، کل فرآیند تصمیم گیری ممکن است به درستی صورت نگیرد (۱۴, ۳۸).

این شامل تمایز، ادغام و تبدیل داده‌ها به شکل قابل درک و کاهش حجم داده‌ها است که باعث بهبود درک از این اطلاعات می شود. در بخش فوق، در مورد تکنیک‌های مختلف داده‌کاوی، قسمت های مختلفی که برای تشخیص بیماری ها استفاده می شود بحث شده است و در این میان، طبقه‌بندی پرکاربردترین الگوریتم است که در مقایسه با سایر الگوریتم ها دارای مزایای فراوانی است و درک بهتری از داده‌ها را به همراه دارد (۳۸).

همچنین مزایا و معایب تکنیک‌های داده‌کاوی در بالا مورد بحث قرار گرفته است. مطالعاتی که در آینده انجام خواهد شد می تواند در این زمینه باشد که انجام پیش‌پردازش به‌عنوان یک روش تشخیصی تکامل یافته است، و به عنوان یک چارچوب مدل‌سازی داده‌های پراکنده پیشنهاد شده است. همچنین در مطالعات بیان شده است که تبدیل مجموعه داده‌ها برای تبدیل ویژگی‌های اسمی به ویژگی‌های عددی انجام می‌شود تا از تولید زیر مجموعه‌ها پشتیبانی کند (۱۴).

یعنی پاسخ‌های مثبت، مانند بله، خوب و دارد، به عنوان ۱ تعریف می‌شوند، در حالی که پاسخ‌های منفی مانند نه، ضعیف و وجود ندارد، به عنوان ۰ تعریف می‌شوند. هدف تولید زیر مجموعه‌هایی از داده‌ها و سپس آموزش دانشجویان با استفاده از زیرمجموعه‌ها و مدل‌های ارائه شده می باشد (۵)، همچنین با ترکیب موارد تولید شده و ارائه مدل های ترکیبی از اطلاعات به دست آمده از بیماران مبتلا به اختلال مزمن کلیوی، می توان آموزش های لازم در زمینه تشخیص و درمان هدفمند این دسته از بیماران را فراهم کرد (۷).

با توجه به اندازه گیری شدت CKD، در حال حاضر میزان فیلتراسیون گلومرولی رایج ترین عامل اندازه گیری مورد استفاده در زمینه مراقبت های بهداشتی برای تخمین عملکرد کلیه است. محاسبه GFR توسط پزشک در زمینه مراقبت‌های بهداشتی از طریق کراتینین خون بیمار، سن، نژاد، جنسیت و عوامل دیگر بسته به نوع فرمول‌های محاسباتی شناخته‌شده رسمی مورد استفاده، انجام می‌شود. GFR چگونگی سلامت کلیه بیمار را در رابطه با موضوع بیماری نشان می‌دهد. پزشک این اقدامات را برای تعیین مراحل مختلف بیماری کلیوی بیمار انجام می‌دهد (۲۱, ۳۵).

بر اساس موارد بیان شده در این بخش اهمیت استفاده از این روش ها در تشخیص زودهنگام و درمان بیماران مبتلا به اختلالات مزمن کلیوی مشخص می گردد، بر این اساس با استفاده از این روش می توان وضعیت این دسته از بیماران را بهبود بخشیده و باعث کاهش علایم و عوارض در آن ها گردید.

نتیجه گیری

هدف این مقاله تحلیل تکنیک‌های مختلف داده‌کاوی در حوزه پزشکی و بر روی بیماران مبتلا به بیماری مزمن کلیوی و برخی از الگوریتم های مورد استفاده برای پیش‌بینی بیماری‌های کلیوی است. از بررسی فوق، ثابت شده است که نتایج ممکن است برای مراحل مختلف تشخیص بیماری کلیوی بر اساس ابزار و تکنیک‌های مورد استفاده متفاوت باشد. داده‌کاوی نتایج بهتری را در تشخیص بیماری در صورت استفاده از تکنیک‌های مناسب ارائه می‌دهد؛ بنابراین، داده‌کاوی زمینه مهمی برای پیش‌بینی بیماری‌های متفاوت مانند اختلالات کلیوی و ارائه مراقبت های بهداشتی فراهم می کند.

محدودیت ها و پیشنهادات

این مطالعه مانند هر مطالعه دیگری با محدودیت هایی روبه رو بوده است، یکی از مهم ترین محدودیت ها عدم دسترسی به اطلاعات برخی مطالعات انجام شده در این زمینه بوده است. همچنین پیشنهاد می شود که مطالعات بیشتری در این زمینه انجام شود و در مطالعات انجام شده در این زمینه وضعیت چندین بیماری بر اساس روش‌های داده‌کاوی مورد ارزیابی و مقایسه قرار بگیرد.

References

1. Patil PM. Review on prediction of chronic kidney disease using data mining techniques. International Journal of Computer Science and Mobile Computing. 2016;5(5):135-41.
2. Chiu RK, Yu-Jing R, editors. Constructing Models for Chronic Kidney Disease Detection and Risk Estimation. Proceedings of 22nd IEEE International Symposium on Intelligent Control, Singapore; 2011.
3. Vijayarani S, Dhayanand S, Phil M. Kidney disease prediction using SVM and ANN algorithms. International Journal of Computing and Business Research (IJCBR). 2015;6(2):1-12.
4. Baby PS, Vital TP. Statistical analysis and predicting kidney diseases using machine learning algorithms. International Journal of Engineering Research and Technology. 2015;4(7):206-10.
5. Arasu SD, Thirumalaiselvi R. Review of chronic kidney disease based on data mining techniques. International Journal of Applied Engineering Research. 2017;12(23):13498-505.
6. Dunham MH. Data mining: Introductory and advanced topics: Pearson Education India; 2006.
7. Dubey A. A classification of ckd cases using multivariate k-means clustering. International Journal of Scientific and Research Publications. 2015;5(8):1-5.
8. NKF K. DOQI Clinical Practice Guidelines for chronic kidney disease: evaluation, classification and stratification. Am J Kidney Dis. 2002;39.(۲)
9. Pavithra N, Shanmugavadivu R. Survey on data mining techniques used in kidney related diseases. International Journal of Modern Computer Science. 2016;4(4):178-82.
10. ۱Kaur G, Sharma A, editors. Predict chronic kidney disease using data mining algorithms in hadoop. 2017 International Conference on Inventive Computing and Informatics (ICICI); 2017: IEEE.
11. ۱Aqlan F, Markle R, Shamsan A, editors. Data mining for chronic kidney disease prediction. IIE Annual Conference Proceedings; 2017: Institute of Industrial and Systems Engineers (IISE).
12. ۱Xia P, Gao K, Xie J, Sun W, Shi M, Li W, et al. Data mining-based analysis of Chinese medicinal herb formulae in chronic kidney disease treatment. Evidence-Based Complementary and Alternative Medicine. 2020;2020.
13. Padmanaban KA, Parthiban G. Applying machine learning techniques for predicting the risk of chronic kidney disease. Indian Journal of Science and Technology. 2016;9(29):1-6.
14. Yadav GS, Gupta R, Pandey M. Comparison of different renal function tests for detecting renal dysfunction in patients with Metabolic Syndrome.
15. Potharaju SP, Sreedevi M. An improved prediction of kidney disease using SMOTE. Indian Journal of Science and Technology. 2016;9(31):1-7.

16. Gharibdousti MS, Azimi K, Hathikal S, Won DH, editors. Prediction of chronic kidney disease using data mining techniques. IIE Annual Conference Proceedings; 2017: Institute of Industrial and Systems Engineers (IISE).
17. Elhoseny M, Shankar K, Uthayakumar J .Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific reports. 2019;9(1):1-14.
18. Saringat Z, Mustapha A, Saedudin RR, Samsudin NA. Comparative analysis of classification algorithms for chronic kidney disease diagnosis. Bulletin of Electrical Engineering and Informatics. 2019;8(4):1496-501.
19. Gunarathne W, Perera K, Kahandawaarachchi K, editors. Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). 2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE); 2017: IEEE.
20. Jongbo OA, Adetunmbi AO, Ogunrinde RB, Badeji-Ajisafe B. Development of an ensemble approach to chronic kidney disease diagnosis. Scientific African. 2020;8:e00456.
21. Kirubha V, Priya SM. Survey on data mining algorithms in disease prediction. Int J Comput Trends Tech. 2016;38(3):24-128.
22. Hannan M, Ansari S, Meza N, Anderson AH, Srivastava A, Waikar S, et al. Risk factors for CKD progression: overview of findings from the CRIC study. Clinical Journal of the American Society of Nephrology. 2021;16(4):648-59.
23. Kazancioğlu R. Risk factors for chronic kidney disease: an update. Kidney international supplements. 2013;3(4):368-71.
24. Shlipak MG, Fried LF, Cushman M, Manolio TA, Peterson D, Stehman-Breen C, et al. Cardiovascular mortality risk in chronic kidney disease: comparison of traditional and novel risk factors. Jama. 2005;293(14):1737-45.
25. Gregg LP, Hedayati SS. Management of traditional cardiovascular risk factors in CKD: what are the data? American Journal of Kidney Diseases. 2018;72(5):728-44.
26. Muntner P, Judd SE, Gao L, Gutiérrez OM, Rizk DV, McClellan W, et al. Cardiovascular risk factors in CKD associate with both ESRD and mortality. Journal of the American Society of Nephrology. 2013;24(7):1159-65.
27. Hsu CW, Yamamoto KT, Henry RK, De Roos AJ, Flynn JT. Prenatal risk factors for childhood CKD. Journal of the American Society of Nephrology. 2014;25(9):2105-11.
28. Inaguma D, Imai E, Takeuchi A, Ohashi Y, Watanabe T, Nitta K, et al. Risk factors for CKD progression in Japanese patients: findings from the Chronic Kidney Disease Japan Cohort (CKD-JAC) study. Clinical and experimental nephrology. 2017;21(3):446-56.

29. González MH, Rolla AL. Data mining techniques applied to statistical prediction of monthly precipitation in Gran Chaco Argentina. Theoretical and Applied Climatology. 2022:1-17.
30. Saha I, Gourisaria MK, Harshvardhan G. Classification System for Prediction of Chronic Kidney Disease Using Data Mining Techniques. Advances in Data and Information Sciences: Springer; 2022. p. 429-43.
31. Lee SJ, Siau K. A review of data mining techniques. Industrial Management & Data Systems. 2001.
32. Qasem MH, Hudaib A, Obeid N, Almaiah MA, Almomani O, Al-Khasawneh A. Multi-agent Systems for Distributed Data Mining Techniques: An Overview. Big Data Intelligence for Smart Applications. 2022:57-92.
33. YAVUZ Ö. A Classification and Clustering Approach Using Data Mining Techniques in Analysing Gastrointestinal Tract. International Scientific and Vocational Studies Journal. 2022;5(2):254-65.
34. Kaur J, Dharni K. Application and performance of data mining techniques in stock market: A review. Intelligent Systems in Accounting, Finance and Management.
35. Firman G. Definition and Stages of Chronic Kidney Disease, Released on Feb. 28. 2009.
36. Navas de Maya B, Arslan O, Akyuz E, Kurt RE, Turan O. Application of data-mining techniques to predict and rank maritime non-conformities in tanker shipping companies using accident inspection reports. Ships and Offshore Structures. 2022;17(3):687-94.
37. Bandgar TR, Kalra S, Sahay M. Metabolic syndrome leading to chronic kidney disease: An emerging threat. Indian journal of endocrinology and metabolism. 2012;16(2):151.
38. Thiyagaraj M, Suseendran G. Review of Chronic Kidney Disease Based on Data Mining Techniques.