

## ساخت مدلی برای تعیین نوع درمان سنگ حالب با راه کار داده کاوی

تاریخ دریافت مقاله: ۱۳۸۸/۰۱/۱۷ تاریخ پذیرش: ۱۳۸۸/۰۵/۲۴

## چکیده

**زمینه و هدف:** داده کاوی یکی از زمینه های چند تخصصی برآمده از زمینه های علمی چون آمار، ریاضی، رایانه و هوش مصنوعی است که کاربردهای آن در حیطه های گوناگون پژوهشی، مدیریتی و اجرایی سلامت و درمان رو به گسترش است. داده کاوی استخراج دانش و اطلاعات ناشناخته، نهان و ارزشمند از داده های موجود در پایگاه های داده به طور خودکار است. به کارگیری داده کاوی بر روی داده های پزشکی، دستاوردهای حیاتی، با ارزش و اثرگذاری را به ارمغان آورده و از جمله می تواند برای یاری رسانی به پزشکان در زمینه تشخیص نوع بیماری و یا انتخاب نوع درمان مناسب کمک شایانی در زمینه نجات جان انسان ها انجام دهد. در این مقاله داده کاوی روی اطلاعات مربوط به بیماران مبتلا به سنگ حالب انجام شده است تا بتوان به کمک آن سیستم کمک تصمیمی طراحی کرد که جهت تعیین نوع درمان بیماران مبتلا به سنگ حالب، با توجه به خصوصیات بیماران استفاده گردد. روش بررسی: اطلاعات مربوط به این بیماران از بیمارستان شهید هاشمی نژاد جمع آوری گردید و پس از انجام یک سری پیش پردازش اولیه، درخت تصمیم مربوط به انتخاب روش درمانی مناسب برای بیماران استخراج گردید. یافته ها: با کمک این روش، پارامترهای موثر روی انتخاب روش درمانی مناسب برای بیماران مبتلا به سنگ حالب استخراج شد. نتیجه گیری: با استفاده از این درخت برای انتخاب روش درمانی موفق برای بیماران مبتلا به سنگ حالب، عملکرد بیمارستان بهبود می یابد یعنی تعداد بیشتری از بیماران، پس از انجام روش درمانی، به طور کامل بهبود می یابند.

**کلمات کلیدی:** سنگ حالب، داده کاوی، انتخاب مشخصه ها و درخت تصمیم

محمد مهدی سپهری\*

پریسا رهنما<sup>۲</sup>پژمان شادپور<sup>۳</sup>بابک تیمورپور<sup>۱</sup>

۱- گروه مهندسی صنایع

۲- گروه مهندسی فناوری اطلاعات

دانشگاه تربیت مدرس

۳- گروه اورولوژی دانشگاه علوم پزشکی ایران

\*نویسنده مسئول، تهران، بخش مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه تربیت مدرس، صندوق پستی ۱۶۱-۱۴۱۱۵ تهران. تلفن: ۸۲۸۸ ۳۳۷۹  
email: mehdi.sepohri@modares.ac.ir

## مقدمه

درک کامل مسئله پیش نیاز ضروری برای انتخاب روش مناسب داده کاوی و کشف دانش مفید از میان داده ها می باشد.<sup>۱، ۲</sup> جمع آوری و پیش پردازش داده ها: پیش پردازش داده ها شامل مراحل است. این مراحل عبارتند از یکپارچه سازی، حذف نویزها، مقابله با مقادیر مفقوده و تغییر شکل داده ها، کاهش تعداد داده ها و یا تعداد ویژگی ها و غیره.<sup>۳</sup> داده کاوی: مرحله سوم همان مرحله داده کاوی می باشد که با انجام آن الگوها و مدل های پنهان در میان داده ها استخراج می گردد. در این مرحله ما باید نخست وظیفه داده کاوی (data mining task) و سپس روش داده کاوی را انتخاب نماییم. ۴- تفسیر و ارزیابی نتایج: مرحله چهارم شامل تفسیر نتایج به دست آمده از مرحله سوم می باشد. لزوماً نتایج به دست آمده از مرحله سوم قابل اطمینان نیستند و ممکن است بیانگر دانش مفید و قابل استفاده

عبارت داده کاوی (Data mining) مترادف با یکی از عبارتهای استخراج دانش، برداشت اطلاعات، واریسی داده ها و حتی لایروبی کردن داده ها است که در حقیقت کشف دانش در پایگاه داده ها (Knowledge Discovery in Databases (KDD) را توصیف می کند. به طور ساده داده کاوی شامل استخراج دانش از حجم عظیم داده ها می شود. البته عباراتی مانند کشف دانش از داده، استخراج دانش، آنالیز الگوها و لایروبی داده را نیز می توان معادل واژه داده کاوی به کار برد.<sup>۱</sup> فرآیند کشف دانش دارای چندین مرحله می باشد که دقت در انجام هر کدام از این مراحل، روی کیفیت کل فرآیند تاثیرگذار است. این مراحل عبارتند از: ۱- فهم و تعریف مسئله: مرحله اول فهم حوزه کاری و مسئله ای است که سعی در پیدا کردن راه حل برای آن داریم.

است که از جمله می توان به این موارد اشاره کرد: Zupan و Lavrac (۲۰۰۵) در فصلی ویژه از یک کتاب به معرفی مبسوط داده کاوی در پزشکی پرداخته اند.<sup>۴</sup> انجمن مهندسان برق و الکترونیک (IEEE) در کنفرانس بین المللی داده کاوی خود در سال ۲۰۰۷، کارگاه تخصصی داده کاوی در پزشکی را با نام (DMMed '07) برگزار نمود. با این حال، در ارتباط با مبحث داده کاوی و بیماری های کلیوی مقالات معدودی منتشر شده اند که در این راستا می توان به مقالات Shah (۲۰۰۳) و Kusiak (۲۰۰۵) که در واقع یک گروه پژوهشی متمرکز بر روی بیماران دیالیزی هستند توجه داشت.<sup>۵</sup> این گروه نشان داده اند که با استفاده از پارامترهای اندازه گیری شده می توانند به پیش بینی طول ادامه حیات بیماران جدید بپردازند. در مورد داده کاوی روی داده های به دست آمده از بیماران مبتلا به سنگ کلیه (حالب) تاکنون فعالیتی صورت نگرفته است. در واقع، انجام این تحقیق از نظر مطالعه روی بیماران مبتلا به سنگ کلیه و اعمال داده کاوی روی ویژگی های این بیماران جهت تعیین نوع درمان مناسب برای آنها، کاری نو می باشد که پس از مشورت با متخصصین اورولوژی و تاکید بر با ارزش بودن نتایج چنین تحقیقی، به عنوان موضوع این کار تحقیقاتی انتخاب گردید. برای انجام داده کاوی روش های مختلفی وجود دارد و بسته به چگونگی نمایش خروجی و نوع داده ها، می توان یکی از این روش ها را جهت استخراج الگوهای نهان در داده ها انتخاب نمود.

سنگ کلیه و روش های درمانی موجود برای آن: بیماری سنگ کلیه، یک بیماری معمول است که درصد قابل توجهی از مردم جهان به آن مبتلا هستند. البته بسیاری از این سنگ ها بدون انجام هیچگونه روش درمانی خود به خود دفع می شوند. سنگ های کلیوی انواع مختلفی دارند ولی ۷۰٪ سنگ ها دارای اجزای کلسیمی هستند. این سنگ ها حاوی فسفات کلسیم و یا اکسالات کلسیم هستند. تعداد سنگ های غیر کلسیمی بسیار کمتر از سنگ های کلسیمی است. سنگ های غیر کلسیمی می توانند از جنس اسید اوریک (Uric acid)، سیستین (Cystine) و یا اجزای شناخته نشده باشند.<sup>۶</sup> درمان سنگ کلیه انواع مختلف دارد که با توجه به شرایط مختلف، یکی از انواع این درمان ها برای بیمار انتخاب می شود. این روش های درمانی شامل Open Surgery و Laparoscopy، PCNL، ESWL، TUL می شود.

۱- transurethral lithotripsy (TUL): این شیوه درمان یک روش سنگ شکنی درون اندامی یا در واقع سنگ شکنی داخل حالب است.

نباشند. به همین خاطر باید این نتایج را بگونه ای ارزیابی نمود. برای ارزیابی مدل به دست آمده می توان در ابتدای امر داده ها را به دو دسته آموزش (train) و آزمایش (test) تقسیم نمود، مدل را روی داده های آموزش ساخت و سپس نتایج به دست آمده را روی داده های تست آزمایش کرد و دقت مدل را محاسبه نمود. ۵- استفاده از دانش کشف شده: مرحله آخر استفاده از دانش استخراج شده از داده ها به صورت عملی می باشد. در حقیقت هدف از انجام مراحل مختلف کشف دانش، دستیابی به نتایجی است که بتوان از آنها در دنیای واقعی و برای بهبود کارایی سازمان ها استفاده کرد. این دانش استخراج شده می تواند به عنوان یک سیستم کمک تصمیم در دنیای واقعی مورد استفاده قرار گیرد. داده کاوی روی داده های پزشکی از اهمیت بالایی برخوردار است و طراحی سیستم های کمک تصمیم جهت یاری رساندن به پزشکان در زمینه تشخیص نوع بیماری و یا انتخاب نوع درمان مناسب، با کمک داده کاوی می تواند کمک شایانی در زمینه نجات جان انسان ها انجام دهد. یکی از اندام هایی که در بدن نقش اساسی و کلیدی ایفا می کند، کلیه می باشد. انجام آزمایشات مناسب و انتخاب روش های درمانی مطلوب برای افرادی که به یکی از بیماری های کلیوی مبتلا هستند، بسیار حائز اهمیت است چراکه کوتاهی در درمان بیماری های کلیوی می تواند منجر به از کارافتادگی سایر ارگان های بدن و در برخی موارد منجر به مرگ بیمار شود. یکی از بیماری های مربوط به کلیه، تشکیل سنگ کلیه در افراد می باشد. امروزه روش های درمانی مختلف برای برخورد با این بیماری وجود دارد اما انتخاب روش درمانی مناسب که علاوه بر آسیب رسانی کمتر به فرد، دارای خروجی مطلوب نیز باشد، کاری پیچیده است. معمولاً روش درمانی برای این بیماری با توجه به خصوصیات سنگ، انتخاب می گردد که البته در بسیاری از موارد روش درمانی مناسب توسط پزشک برای بیمار انتخاب نمی شود. بدین جهت طراحی یک سیستم تصمیم یار جهت انتخاب نوع درمان سنگ کلیه با توجه به خصوصیات سنگ و بیمار و با کمک راهکار داده کاوی، می تواند به پزشکان در انتخاب درمان یاری رساند.

جدید بودن موضوع تحقیق و لزوم انجام مطالعه: همزمان با گسترش و رشد داده کاوی و کاربردهای آن در دهه اخیر، توجه پژوهشگرانی از حوزه پزشکی و سلامت را به خود جلب نموده و به کارگیری آن در تحقیقات این حوزه روندی رو به تزاید پیدا کرده

شاخه‌های درخت با توجه به مقدار آن ویژگی پایین می‌آییم. سپس این فرآیند با امتحان گره بعدی که در انتهای شاخه‌ی انتخاب شده قرار دارد، ادامه می‌یابد تا نهایتاً به یک برگ برسیم. تاکنون از درخت تصمیم به عنوان روش داده‌کاوی جهت استخراج الگوهای نهان در داده‌های پزشکی، استفاده‌های زیادی شده است.

طراحی تحقیق، ساخت مدل: در این تحقیق نخست فاکتورهای موثر روی نوع درمان انتخابی برای بیماران مبتلا به سنگ حالب استخراج شد. سپس اطلاعات مورد نیاز از بیمارستان شهید هاشمی‌نژاد تهران جمع‌آوری شد. پس از آن یک‌سری کارهای پیش‌پردازشی جهت آماده‌سازی داده‌ها، برای ساخت مدل، صورت پذیرفت. پس از آن درخت تصمیم مربوط به نمونه‌ها به کمک الگوریتم CART ساخته شد و در نهایت کارایی مدل ساخته شده مورد ارزیابی قرار گرفت. برای انجام داده‌کاوی و کارهای پیش‌پردازشی مربوط به آن، نرم‌افزار (http://www.r-project.org) R مورد استفاده قرار گرفت.

تهیه check list جهت جمع‌آوری اطلاعات مربوط به بیماران: پس از مطالعه روی بیماران مبتلا به سنگ حالب و عوامل موثر روی تعیین نوع درمان آن‌ها، یک‌سری از ویژگی‌ها استخراج گردید که این ویژگی‌ها در جدول ۱ آورده شده است: در مورد نوع بیهوشی، سه نوع بیهوشی در نظر گرفته شد: بیهوشی عمومی، بیهوشی regional، بیهوشی local. در مورد ویژگی output، طبق نظر متخصص سه حالت در نظر گرفته شد، Stone free: پس از انجام روش درمانی سنگ به کلی از بین رفته و درمان شده است. Break and remain: با انجام روش درمانی سنگ به قطعات کوچک‌تری شکسته شده است ولی این قطعات هنوز در کلیه بیمار باقی مانده‌اند. Fail: حتی با انجام روش درمانی نیز سنگ بدون کوچک‌ترین تغییری باقی مانده است. در این حالات اصطلاحاً گفته می‌شود که درمان شکست خورده است. پیش‌پردازش داده‌ها: انجام پیش‌پردازش روی داده‌ها یکی از مراحل مهم در این امر می‌باشد. نمونه‌های موجود نیز نیازمند یک‌سری پردازش‌های اولیه بودند تا برای ساخت مدل نهایی آماده گردند. مرحله اول پیش‌پردازش حذف نمونه‌هایی بود که در جامعه آماری مورد مطالعه قرار نمی‌گرفتند. این نمونه‌ها آنهایی بودند که مکان سنگ در آنها در جایی غیر از کلیه و میزنای، برای مثال مثانه ذکر شده بود. برای برخورد با مقادیر مفقوده، از محاسبه میزان شباهت

این روش عموماً موفقیت‌آمیز است که در آن بیمار حداقل یک روز بستری شده و با استفاده از سوند حالب سنگ شکنی انجام می‌گیرد. ۲- Extracorporeal Shock Wave Lithotripsy (ESWL): در روش (ESWL) یا (SWL) امواج لرزاننده (shock waves) با انرژی بالا را که توسط یک منبعی خارج از بدن ایجاد شده‌اند، روی سنگ متمرکز می‌کند و از این انرژی برای خرد کردن سنگ استفاده می‌کند. قطعات باقی‌مانده تا حدی کوچک هستند که توسط ادرار از بدن دفع می‌شوند. ۳- Percutaneous Nephrolithotomy (PCNL): سنگ‌های بزرگ را می‌توان به کمک جراحی از کلیه خارج نمود. جراح یک برش کوچک در پشت بیمار ایجاد می‌کند و سپس از ابزاری که نفروسکوپ (nephroscope) نامیده می‌شود استفاده کرده و به کمک آن یا سنگ را خارج می‌کند و یا با استفاده از امواج یا لیزر سنگ را می‌شکند. ۴- ureteroscopy: این نوع درمان زمانی انتخاب می‌شود که سنگ در میزنای قرار داشته باشد. برای درمان، متخصص کلیه یک وسیله را که ureteroscope نامیده می‌شود از طریق پیش‌آبراه (urethra) وارد مثانه و سپس میزنای می‌کند. سپس سنگ به وسیله امواج اولتراسوند خارج می‌شود. ۵- جراحی باز: در این نوع درمان بیمار برای رفتن به اتاق عمل آماده می‌شود. سپس جراح پوست را باز می‌کند و روی لگنچه و یا میزنای شکافی ایجاد می‌کند به طوری که بتواند به سنگ دسترسی پیدا کند.

## روش بررسی

این مطالعه از آبان ماه سال ۸۵ تا آذر سال ۸۶ انجام شده است. اطلاعات مربوط به بیماران از بیمارستان شهید هاشمی‌نژاد جمع‌آوری شده و کارهای مربوط به ساخت مدل (ورود داده‌ها، پیش‌پردازش و ساخت مدل) در آزمایشگاه فناوری اطلاعات دانشگاه تربیت مدرس انجام شده است.

درخت تصمیم (Decision Tree): یکی از روش‌های داده‌کاوی روی پایگاه‌های داده، ساخت درخت تصمیم مربوط به داده‌ها می‌باشد. درخت تصمیم یک ساختار درخت مانند دارد، در حالی که هر گره ورودی نشان‌دهنده یک امتحان روی یک ویژگی، هر شاخه خارج شده از آن گره بیان‌کننده یکی از مقادیر ممکن برای همان ویژگی و برگ‌ها نشان‌دهنده کلاس‌ها هستند. فرآیند دسته‌بندی هر نمونه با امتحان ویژگی بیان شده در گره ریشه شروع شده و سپس از

نمونه‌ها در نظر گرفته شد (stone free و break and remain و fail). تعداد این کلاس‌ها در بین نمونه‌های موجود غیر متوازن بود. در ساخت درخت تصمیم برای رسیدن به نتایج معتبرتر باید متوازن‌سازی دسته‌ها صورت پذیرد به طوری که تعداد نمونه‌های دسته بیشتر حداکثر دو برابر نمونه‌های سایر دسته‌ها باشد. بدین منظور قبل از انجام داده‌کاوی متعادل‌سازی نمونه‌ها با نمونه‌گیری تصادفی از کلاس‌های با تعداد کمتر به میزان نصف کلاس بیشتر انجام گرفت.<sup>۱۴</sup>

انتخاب مشخصه‌ها: عامل مهم در ساخت مدلی با صحت بالا، انتخاب درست ویژگی‌ها می‌باشد. ویژگی‌های موثر در ساخت یک درخت تصمیم، باید وابستگی کمی با یکدیگر داشته باشند. در این تحقیق برای انتخاب ویژگی‌های مناسب از بین ویژگی‌های موجود feature selection انجام شد. دو روش برای feature selection وجود دارد: (۱) filter algorithms و (۲) Wrapper algorithms. در الگوریتم‌های Filter روش ارزیابی زیر گروه ویژگی‌های انتخاب شده، مستقل از روش انتخابی برای انجام داده‌کاوی می‌باشد. ولی در الگوریتم wrapper برای بررسی زیر گروه ویژگی‌های انتخاب شده، از خود روش استنتاجی در اینجا درخت تصمیم به عنوان تابع ارزیابی استفاده می‌شود.<sup>۱۵</sup> هنگامی که تعداد متغیرها و نمونه‌ها محدود است، الگوریتم Wrapper می‌تواند ویژگی‌های مناسب جهت ساخت مدل را انتخاب نماید. برای انجام الگوریتم wrapper روش‌های مختلفی وجود دارد. یکی از این روش‌ها Random Forest نامیده می‌شود. روش استفاده شده برای انتخاب ویژگی‌ها در این تحقیق، Random Forest می‌باشد. جهت بالابردن صحت کار، انتخاب ویژگی‌ها توسط الگوریتم random forest، ۱۰ بار انجام شد و ویژگی‌های انتخاب شده در هر مرحله در یک data frame ذخیره شد. پس از آن با استفاده از پکیج arules، ویژگی‌هایی که در تمامی ۱۰ بار اجرای random forest انتخاب شده بودند، تعیین گردیدند تا از این ویژگی‌ها در ساخت درخت استفاده شود. پس از اجرای Random Forest روی ویژگی‌های، تعدادی از آنها به عنوان ویژگی‌های نهایی جهت ساخت مدل انتخاب گردید. نام این متغیرها در جدول ۲ آورده شده است.

جدول ۲: ویژگی‌های انتخابی توسط الگوریتم Random Forest

حالب
اندازه
سن
درمان

بین نمونه‌ها استفاده شد. برای محاسبه میزان شباهت بین دو متغیر از دستور daisy استفاده گردید. فرض کنید  $X_{ij}$  مقدار مشاهده شده متغیر  $j$  ام از نمونه  $i$  ام باشد. مسئله در اینجا این است که بینیم دو نمونه  $s$  ام و  $t$  ام چقدر به هم شبیه هستند. فرض بر این است که دو نمونه به صورت  $X_t = (x_{t1} \dots x_{tp})$  و  $X_s = (x_{s1} \dots x_{sp})$  برای مثال برای بررسی میزان شباهت دو نمونه می‌توان فاصله اقلیدسی بین دو آنها را محاسبه نمود. طریقه محاسبه این فاصله برای دو نمونه  $X_t$  و  $X_s$  در این رابطه آورده شده است:

$$d(s, t) = \left( \sum_{j=1}^p (x_{sj} - x_{tj})^2 \right)^{\frac{1}{2}}$$

هر چقدر میزان این فاصله بین دو نمونه کمتر باشد، آن دو نمونه بیشتر به یکدیگر شبیه هستند. پس از محاسبه فاصله اقلیدسی بین نمونه دارای متغیر مفقوده، ۱۰ نمونه شبیه‌تر به آن انتخاب شده و میانه متغیر عددی آن ۱۰ نمونه، به عنوان جایگزین متغیر مفقوده در نظر گرفته می‌شود. مرحله بعدی پیش‌پردازش تبدیل سال تولد بیماران به سن بود. بدین ترتیب سال تولد تمامی بیماران از عدد ۱۳۸۶ کسر گردید تا سن افراد به دست آید. مرحله چهارم از پیش‌پردازش شامل حذف ویژگی‌هایی بود که در نتایج تأثیری نداشتند. این متغیرها شامل ID و شماره سریال مربوط به بیماران می‌شد.

متوازن‌سازی نمونه‌ها: در این تحقیق سه دسته (برچسب کلاس) برای

جدول ۱: نام و نوع مشخصه‌های جمع‌آوری شده از بیماران

نام مشخصه	نوع مشخصه
جنسیت	دسته‌ای (categorical) (زن، مرد) <sup>۷</sup>
سن	عددی (numeric) <sup>۷</sup>
سابقه دفع سنگ	دسته‌ای (بله، خیر) <sup>۸</sup>
سابقه سنگ شکنی	دسته‌ای (بله، خیر) <sup>۹</sup>
سابقه عمل جراحی	دسته‌ای (بله، خیر) <sup>۹</sup>
اندازه	عددی <sup>۱۰، ۱۱</sup>
تعداد سنگ‌ها	عددی
مکان	دسته‌ای (pelvis, low calyx, high calyx, mid calyx, high ureter, low ureter) <sup>۱۲</sup>
هیدرونفروز	دسته‌ای (بله، خیر)
ناشفافی (opacity)	دسته‌ای (none, semi low, semi high, opac)
بیهوشی	ترتیبی <sup>۱-۳</sup>
نوع درمان	دسته‌ای (PCNL, TUL, laparoscopy, SWL, open surgery)
خروجی	دسته‌ای (از بین رفتن سنگ (Stone free)، شکستن و باقی ماندن سنگ در کلیه (Break and remove)، شکست خوردن درمان (fail))
اندازه سنگ باقی مانده در کلیه پس از درمان (sizeRS)	عددی

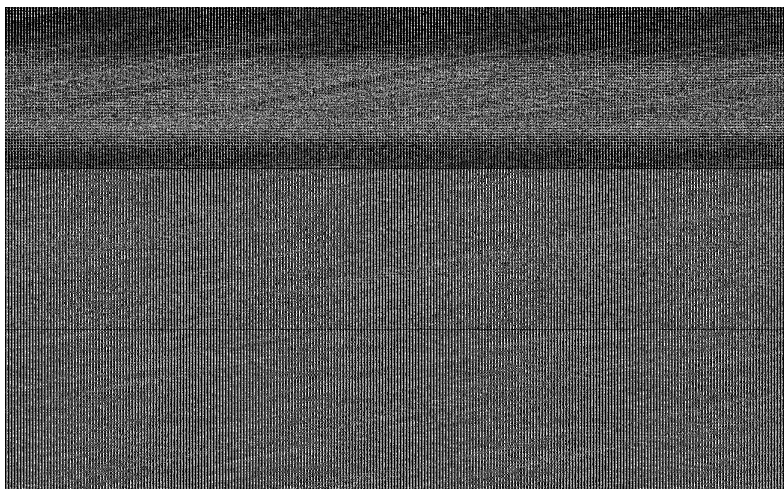
درمانی stone free می‌شدند. در بیمارستان الگوریتمی وجود دارد که در آن با توجه به مکان سنگ و اندازه سنگ، یک نوع روش درمانی پیشنهاد می‌شود. البته پارامتر سومی نیز وجود دارد که در انتخاب روش درمانی موثر است. این پارامتر به impacted و یا non-impacted بودن سنگ اشاره دارد. از آنجایی که تعیین این ویژگی مشکل می‌باشد و کاملاً بستگی به نظر پزشک دارد، آن را در بین ویژگی‌های موجود در check list تهیه شده قرار ندادیم. به همین علت برای بررسی میزان عملکرد الگوریتم بیمارستانی مجبور شدیم این ویژگی را حذف کرده و فقط مکان و اندازه سنگ را در نظر بگیریم. الگوریتم مورد استفاده در بیمارستان در شکل ۱ نشان داده شده است. در این مرحله باید عملکرد الگوریتم بیمارستانی مورد بررسی قرار می‌گرفت. منظور از عملکرد تعداد نمونه‌هایی است که اگر طبق الگوریتم بیمارستانی نوع درمانی خاص برای آنها تعیین گردد، stone free خواهند شد. پس از بررسی مشخص گردید که میزان عملکرد الگوریتم درختی بیمارستان ۷۷٪ می‌باشد.

میزان بهبود عملکرد: پس از محاسبه میزان عملکرد الگوریتم بیمارستان و مدل ساخته شده از سیستم ذهنی پزشکان، نوبت به بررسی عملکرد مدل ساخته شده (KSDT) بود. هدف از ساخت این مدل افزایش درصد نمونه‌هایی است که پس از انجام روش درمانی stone free می‌شوند. حال برای نمونه‌های موجود علاوه بر درمان واقعی انجام شده روی آنها، چهار نوع درمان دیگر نیز در نظر گرفته

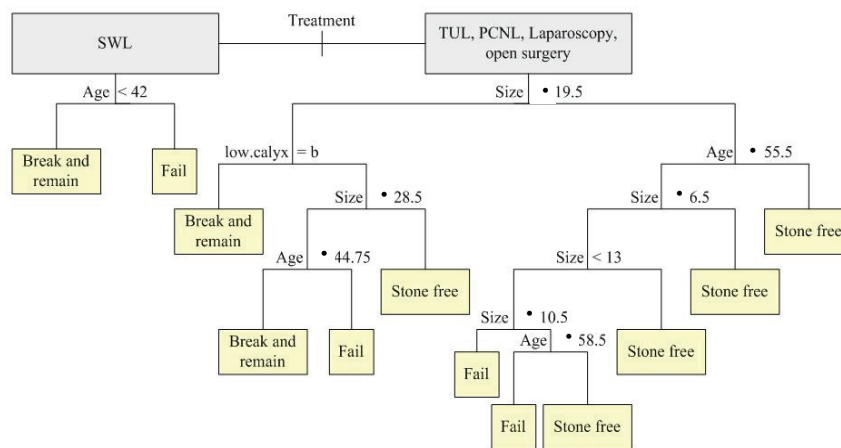
ساخت درخت تصمیم: برای ساخت مدل جهت تعیین نوع درمان بیماران مبتلا به سنگ کلیه از درخت تصمیم استفاده شد. از آنجایی که در بیمارستان شهید هاشمی نژاد یک الگوریتم درخت مانند وجود دارد که تعیین نوع درمان براساس آن انجام می‌شود، ساخت مدلی که خروجی آن به صورت یک درخت نمایش داده شود می‌تواند به خوبی در بیمارستان مورد استفاده قرار گیرد. در ساخت درخت تصمیم از متد one-level-out cross validation استفاده گردید. در این روش در هر بار ساخت درخت یک نمونه به عنوان نمونه test کنار گذاشته شده و مدل روی بقیه نمونه‌ها ساخته می‌شود و صحت آن روی نمونه test محاسبه می‌گردد. اگر K نمونه در پایگاه داده موجود باشد، مدل K بار ساخته شده و صحت نهایی از میانگین‌گیری روی K صحت به دست آمده در هر بار ساخت مدل، محاسبه می‌شود. صحت در اینجا به عنوان کسری از نمونه‌ها که توسط درخت دسته‌بندی شده‌اند، تعریف می‌شود. در این تحقیق در نهایت میانگین‌گیری روی این K صحت صورت پذیرفت تا صحت نهایی درخت به دست آید. پس از انجام این مراحل صحت قابل قبولی حدود ۸۵٪ به دست آمد.

## یافته‌ها

مقایسه نتایج: پس از ساخت درخت حال نوبت مقایسه عملکرد مدل با الگوریتم موجود در بیمارستان بود. در حقیقت هدف از ساخت این مدل افزایش درصد نمونه‌هایی بود که پس از انجام روش



شکل - ۱: الگوریتم مورد استفاده در بیمارستان هاشمی نژاد جهت انتخاب نوع درمان سنگ کلیه (تهیه شده توسط دکتر پژمان شادپور قبل از انجام این مطالعه تحقیقاتی)



شکل - ۲: درخت تصمیم جهت تعیین نوع درمان و خروجی درمان برای بیماران مبتلا به سنگ کلیه

بیمار، پنج نوع درمان موجود در نظر گرفته شده و درصد موفقیت هر کدام از این درمان‌ها پیش‌بینی می‌شود. پزشک می‌تواند با توجه به میزان موفقیت هر کدام از درمان‌ها، درمان مناسب را برای بیمار انتخاب نماید. در الگوریتم بیمارستان، برای انتخاب نوع درمان فقط دو ویژگی اندازه و مکان سنگ مهم است، ولی در KSDT سن بیمار نیز در تصمیم‌گیری موثر است. همچنین در الگوریتم بیمارستان برای سنگ‌های زیر ۱۰ و بالای ۱۰ میلی‌متر دو روش درمانی در نظر گرفته می‌شود، در صورتی که با توجه به مدل ساخته شده باید برای انتخاب نوع درمان، سنگ‌های بیشتر و کمتر از هفت میلی‌متر، بیشتر و کمتر از ۱۱ میلی‌متر، بیشتر و کمتر از ۱۳ میلی‌متر، بیشتر و کمتر از ۲۰ و بیشتر و کمتر از ۲۸ میلی‌متر را در نظر گرفت و با توجه به این اندازه نوع درمان را تعیین نمود.

شد. یعنی برای هر نمونه هر پنج نوع درمان را در نظر گرفتیم و به ازای هر نوع درمان با کمک KSDT خروجی درمان تعیین کردیم. نمونه‌هایی که حداقل به ازای یکی از این پنج نوع درمان stone free می‌شدند، شمارش شدند. تعداد این نمونه‌ها تقسیم بر کل نمونه‌ها، میزان عملکرد مدل را بازگو می‌کند. این مقدار حدود ۸۵٪ است. درخت ساخته شده در این تحقیق در شکل ۲ آورده شده است.

## بحث

از آنجایی که عملکرد این مدل، بهتر از عملکرد الگوریتم بیمارستانی و مدل ذهنی پزشکان می‌باشد، از آن می‌توان به عنوان سیستم کمک تصمیم‌گیری جهت تعیین نوع درمان سنگ‌های حالب در بیمارستان استفاده کرد. با کمک این مدل و با توجه به ویژگی‌های

## References

- Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. 2<sup>nd</sup> ed. San Francisco, CA: Morgan Kaufmann Publishers; 2006.
- Ayre LB. Data Mining for Information Professionals, 2006.
- Rahm E, Do HH. Data Cleaning: Problems and Current Approaches. IEEE Bulletin of the Technical Committee on Data Engineering, 2000.
- Lavrac N. Selected techniques for data mining in medicine. *Artif Intell Med* 1999; 16(1): 3-23.
- Shah S, Kusiak A, Dixon B. Data mining in predicting survival of kidney dialysis patients-invariant object approach [Online]. Available from: URL:<http://www.icaen.uiowa.edu/~ankusiak/Res-in-Prog/Dialysis.pdf>
- Kusiak A, Dixon B, Shah S. Predicting survival time for kidney dialysis patients: a data mining approach. *Comput Biol Med* 2005; 35(4): 311-27.
- Chang S. Kidney health: what you need to know about kidney stones [Online]. 2001 Apr 16. Available from: URL:<http://drgreene.healthology.com/kidney-health/article898.htm>
- Ajayi L, Jaeger P, Robertson W, Unwin R. Renal stone disease. *Medicine* 2007; 35(8): 415-9.
- Knoll T. Stone disease. *Eur Urol Suppl* 2007; 6(12): 712-22.
- Argyropoulos AN, Tolley DA. Optimizing shock wave lithotripsy in the 21st century. *Eur Urol* 2007; 52(2): 344-52.
- Schulman C, Chapple C. New Horizons in Urology: Applying New Insights and Technology into Clinical Practice: Introduction and conclusions. *Eur Urol Suppl* 2007; 6(12): 695-700
- Acar C, Küpeli B, Gürocak S, Alkibay T, Güneri C, Ozkan S, et al. Is pelvicaliceal anatomy a risk factor for stone formation in patients with solitary upper caliceal stone? *Urology* 2006; 67(6): 1159-63.
- Basiria A, Ziaeea S, Hosseini Moghaddama S, Hosseini SA, Danesha A, Sharifi Aghdasa F. MP-05.06: The effect of impaction on the outcome of treatment of pelvic kidney stones: Comparing shock wave lithotripsy and percutaneous nephrolithotomy. *Urology* 2006; 68(suppl 1): 80-1.
- Torgo L. Data Mining with R: learning by case studies [Online]. 2002 Nov 14. Available from: URL:<http://www.liaad.up.pt/~ltorgo/DataMiningWithR>
- Liu Y, Schumann M. Data mining feature selection for credit scoring models. *J Operational Res Soc* 2005; 1-10.

## A data mining based model for selecting type of treatment for kidney stone patients

Received: April 06, 2009 Accepted: August 15, 2009

### Abstract

Sepehri M.M.<sup>1\*</sup>  
Rahnama P.<sup>2</sup>  
Shadpour P.<sup>3</sup>  
Teimourpour B.<sup>1</sup>

1- Department of Industrial  
Engineering

2- Group of Information  
Technology Engineering

School of Engineering, Tarbiat  
Modares University

3- Department of Urology, Iran  
University of Medical Sciences

**Background:** Data mining as a multidisciplinary field is rooted in the fields such as statistics, mathematics, computer science and artificial intelligence and has been gaining momentum in scientific, managerial, and executive applications in health care. Data mining can be defined as the automated extraction of valuable, practical and hidden knowledge and information from large data. Applying data mining in medical records and data is of utmost importance for health care givers and providers and brings vital and valuable outcomes. Data mining can help doctors come up with better recommendations and plans for treatment which actually in many respects have significant impact on patients' life and satisfaction. In this paper we have proposed and utilized data mining methods to extract hidden information in medical records of pelvis stone patients with ureteral stone. We have tried to design a decision support system model to be applicable for selecting type of treatment for these groups of patients.

**Methods:** We gathered needed information from Shahid Hashemi Nejad hospital. In this research we have used decision tree as a data mining tool, for selecting suitable treatment for patients with ureteral stone. This model can predict probability of success of each treatment.

**Results:** In this research we extracted effective attributes in selecting type of treatment for patients with ureteral stone.

**Conclusions:** By using this model we can have eight percent improvement in number of patients who have stone free output after treating. In fact, this model has a better functionality than expert system of hospitals.

**Keywords:** Data mining, ureteral stone, feature selection, decision tree

\*Corresponding author: Department of  
Industrial Engineering, School of  
Engineering, Tarbiat Modares University  
P.O. Box: 14115-161, Tehran, Iran.  
Tel: +98-21-8288 3379  
email: mehdi.sepehri@modares.ac.ir