



## ارائه مدلی برای پیش‌بینی نوع صافی همودیالیز با تکنیک‌های داده کاوی

مریم عاشوری<sup>۱</sup>

چکیده

**مقدمه:** دیالیز ناکافی به عنوان یک عامل خطر مرگ و میر برای بیماران کلیوی، ضرورت وجود الگویی برای کمک به پرسنل بخش دیالیز جهت ارائه خدمات درست به بیماران دیالیزی و مدیریت صحیح درمان آن‌ها را نشان می‌دهد. از آنجا که نقش نوع صافی در کفایت دیالیز تعیین کننده می‌باشد، پژوهش حاضر با هدف شناسایی الگویی برای تشخیص نوع صافی همودیالیز صورت گرفته است.

**روش کار:** پژوهش حاضر به روش توصیفی-مقطعی صورت گرفته است. جامعه‌ی پژوهش متشکل از داده‌های بیمارستان علی ابن ابی‌طالب(ع) زاهدان در خردادماه ۹۵ بود. جهت تحلیل داده‌ها از نرم افزار کلمنتاین ۱۲ استفاده شده است. برای مدل‌سازی از الگوریتم‌های درخت تصمیم سی فایو، تقسیم و رگرسیون، آشکارساز تعامل خودکار مجذور کای، درخت آماری کارا و بی‌طرف سریع و شبکه عصبی استفاده گردید.

**یافته‌ها:** مقدار صحت به دست آمده از اجرای الگوریتم‌های درخت تصمیم سی فایو، تقسیم و رگرسیون، آشکارساز تعامل خودکار مجذور کای، درخت آماری کارا و بی‌طرف سریع و شبکه عصبی به ترتیب  $0/9263$ ،  $0/9047$ ،  $0/8872$  و  $0/8720$  می‌باشد. مقادیر به دست آمده برای شاخص‌های حساسیت، شفافیت، صحت، دقت، ارزش اخباری منفی، شاخص F، میانگین هندسی، نرخ مثبت غلط، نرخ منفی غلط، نرخ کشف غلط و نرخ خطا برای مدل درخت تصمیم سی فایو نشان-دهنده عملکرد بهتر این الگوریتم نسبت به سایر الگوریتم‌ها بود.

**نتیجه‌گیری:** قوانین استخراج شده برای یک نمونه جدید با ویژگی‌های مشخص، می‌توانند صافی مناسب برای دیالیز را پیش‌بینی نمایند. از این رو مدل ارائه شده به پیش‌بینی دقیق‌تر نوع صافی همودیالیز و نیز مدیریت صحیح درمان بیماران کمک می‌نماید که این منجر به ارتقای عملکرد سازمان‌های سلامت می‌گردد.

**کلیدواژه‌ها:** صافی همودیالیز، داده کاوی، درخت تصمیم، شبکه عصبی

• وصول مقاله: ۹۵/۰۴/۳۱ • اصلاح نهایی: ۹۵/۰۸/۱۵ • پذیرش نهایی: ۹۵/۱۱/۱۷

## مقدمه

در آستانه قرن ۲۱، برجسته‌ترین واقعه‌ای که جوامع و کادر بهداشتی با آن مواجه است، شیوع روز افزون بیماری‌های مزمن می‌باشد. یکی از این بیماری‌ها که تنش‌های قابل ملاحظه‌ای برای بیمار به همراه دارد، مرحله انتهایی بیماری کلیه (End stage renal disease: ESRD) می‌باشد [۱]. مرحله انتهایی بیماری کلیه یک بیماری وخیم با عواقب سلامتی جدی است [۲] که در آن کلیه‌ها قادر به پاک‌سازی متابولیت‌ها از خون یا حذف آب از بدن نیست. بدون مداخله پزشکی، ممکن است بیماران بمیرند یا در واحد مراقبت‌های ویژه (Intensive Care Units: ICU) برای مدت طولانی بستری گردند [۳]. بروز کلی این بیماری ۲۶۰ مورد در هر یک میلیون نفر جمعیت در سال است و هر ساله تقریباً شش درصد افزایش می‌یابد [۲ و ۳] به طوری که تعداد این بیماران در جهان از ۲۰ نفر در دهه ۷۰ به ۷۵۰ نفر به ازای هر یک میلیون نفر جمعیت در عصر حاضر افزایش یافته است. بر اساس آمار موجود در ایران، سالیانه ۱۲۰ تا ۱۶۰۰ نفر به این بیماری مبتلا می‌شوند [۴] لذا، کنترل و درمان آن، از اهمیت بسزایی برخوردار است [۲].

همودیالیز، رایج‌ترین شیوه درمانی در روش‌های جایگزین در بیماران با نارسایی کلیه است که در ایران و جهان از آن استفاده می‌شود [۲]. بیش از یک میلیون نفر در جهان از طریق دیالیز به حیات خود ادامه می‌دهند. با دسترسی وسیع به دیالیز، زندگی صدها هزار نفر از بیماران مبتلا به مرحله انتهایی بیماری کلیه، طولانی شده است [۱]. اساس همودیالیز بر خارج ساختن مواد زائد از طریق عبور خون از غشای نیمه تراوا است. پاک‌سازی این مواد بر اساس مکانیسم‌های انتشار و اولترافیلتراسیون از صافی است. صافی‌های مورد استفاده و موجود در بازار به دو دسته طبقه‌بندی می‌شوند. صافی لوفلاکس (Low Flux) به عنوان یک فن همودیالیز استاندارد مورد استفاده است که اساس آن استفاده از دیالیزورهای با نفوذپذیری آبی کم است. صافی‌های فلاکس (High Flux)

اشاره به غشاهای غیرسلولزی با نفوذپذیری افزایش یافته در مقایسه با صافی لوفلاکس دارد [۲].

اگر چه شیوه‌های همودیالیز موجود از کارایی مناسبی برخوردار هستند، اما عدم کفایت دیالیز از جمله عوامل مهم در افزایش مرگ و میر این بیماران محسوب می‌گردد [۲ و ۳] که نقش نوع صافی تعیین کننده می‌باشد. عوامل زیادی از جمله زمان دیالیز، سرعت مایع دیالیز، استفاده از دیالیزور با سطح بالا و سرعت جریان خون بر کفایت دیالیز تاثیر می‌گذارد [۴]. در صورت انجام همودیالیز نا کارآمد، بیمار نیازمند افزایش دفعات یا مدت زمان دیالیز خواهد بود که علاوه بر تحمیل هزینه‌های درمانی اضافی بر سیستم بهداشتی درمانی کشور، خطرات ناشی از انتقال عوامل عفونی مهلک مانند ویروس هپاتیت بی، سی و اچ آی وی و سایر آلودگی‌های خونی را دامن گیر بیمار خواهد کرد. این مسائل لزوم کارآمدتر نمودن همودیالیز را مشخص می‌نماید تا در افزایش کیفیت و طول عمر بیماران همودیالیزی موثر باشد [۵ و ۶].

با توجه به این که استفاده از صافی‌های متناسب با جنس فرد در ارتقای کفایت دیالیز تاثیر گذار است [۵]، دو مساله در خصوص نقش نوع صافی در دیالیز مطرح می‌گردد؛ اول اینکه انتخاب صافی با مشخصات پایین تر از صافی واقعی مورد نیاز سبب می‌شود تا متابولیت‌ها از خون و آب از بدن به خوبی حذف نشود [۲] و دوم اینکه طبق نظر متخصصان حوزه دیالیز، انتخاب صافی با مشخصات بالاتر سبب افت فشار خون بیمار و مصرف آب فراوان توسط وی می‌گردد که در هر دو حالت به افزایش ساعت دیالیز یا تعداد دفعات دیالیز نیاز است. مسلم و همکاران برای بررسی کفایت دیالیز و ارتباط آن با نوع صافی، متغیرهای اوهره قبل و بعد دیالیز، مدت زمان دیالیز، اولترافیلتراسیون و وزن قبل و بعد دیالیز را بررسی نمودند. یافته‌ها نشان داد که ارتباط آماری معناداری بین کفایت دیالیز و نوع صافی و نیز بین میزان فسفر و نوع صافی وجود نداشت، اما ارتباط آماری معناداری بین میزان هموگلوبین و نوع صافی وجود داشت [۱]. واحد پرست و روانی پور استفاده از صافی‌های لوفلاکس را یکی از عوامل پایین بودن کفایت دیالیز

برشمردند. پژوهشگران پیشنهاد کردند به علت عدم کفایت دیالیز بیماران همودیالیزی، گردش خون مجدد، سرعت پمپ-های خون و دیالیزیت، نوع صافی و میزان آگاهی پرسنل مورد بررسی قرار گیرد زیرا مشکلات مختلفی چون میزان جریان خون، نوع صافی و نمونه گیری اشتباه می تواند منجر به کاهش کفایت دیالیز در بیماران استیبل (stable) همودیالیزی گردد [6]. برزو و همکاران افزایش جریان خون را به منظور افزایش میزان کفایت دیالیز استنباط نمودند. در این پژوهش توجه به عواملی چون تحمل بیمار، تغییرات همودینامیک بیمار، استفاده از صافی های متناسب با وزن بیمار و سرعت مناسب جریان خون ضروری اعلام شده است [7].

موسسات بهداشتی، معمولاً حجم زیادی از اطلاعات بالینی را درباره وضعیت بیماران، اقدامات پزشکی (درمان، جراحی) و فرآیندهای مراقبت سلامت (پذیرش، ترخیص، درخواست آزمایش) جمع آوری می نمایند. با وجود فراوانی این داده ها، استفاده عملی از این داده ها محدود به روال های بازپرداخت و حسابداری و گاهی مطالعات امراض مسری است [8]. دانشی که در این پایگاه ها وجود دارد، معمولاً نهان و استفاده نشده است [9 و 10] که استخراج آن با استفاده از داده کاوی می تواند به شناسایی قوانین حاکم بر ایجاد، رشد و تسریع بیماری ها منجر شود و اطلاعات ارزشمندی را به منظور شناسایی علل رخداد بیماری ها، پیش بینی و درمان بیماری ها با توجه به عوامل محیطی حاکم در اختیار متخصصین و دست اندرکاران حوزه سلامت قرار دهد [9 و 11]. بنابراین، فنون داده کاوی می تواند به عنوان ابزاری مهم برای استخراج دانش از داده های پزشکی و سلامت مورد استفاده قرار گیرد [10] و طراحی سیستم های تصمیم یار جهت یاری رساندن به پزشکان در زمینه تشخیص نوع بیماری و یا انتخاب درمان مناسب با کمک داده کاوی، می تواند در زمینه نجات جان انسان ها کمک شایانی انجام دهد [12].

### روش کار

مطالعه حاضر از نوع توصیفی-مقطعی است و مجموعه داده های آن متعلق به بیمارستان علی ابن ابی طالب (ع) زاهدان است. جامعه ی پژوهش 409 نفر مراجعه کننده در خرداد 95 به بیمارستان مذکور بود که با مراجعه مستقیم پژوهش گر به بیمارستان و به صورت فایل اکسل تهیه گردید و محتوای داده ها مورد تایید متخصصان حوزه مربوطه قرار گرفت. به منظور رعایت ملاحظات اخلاقی، نوع مطالعه و هدف آن برای پرسنل بخش دیالیز توضیح و به آن ها اطمینان داده شد که داده های گردآوری شده به صورت محرمانه محفوظ خواهد ماند. به منظور آماده سازی داده ها، رکوردهایی که دارای

در ارتباط با مبحث داده کاوی و بیماری کلیوی، مقالات متعددی منتشر شده اند. فرنلدز و همکاران از شبکه عصبی برای پیش بینی مقدار عددی کفایت دیالیز در طول همودیالیز استفاده نمودند و نشان دادند که اگر شبکه عصبی توانایی

مقادیر از دست رفته (Missing Values) بودند یا هیچ وجه تشابهی با سایر داده‌ها نداشتند (Outlier)، حذف گردیدند. در این مطالعه رکوردهایی که مقدار صفر برای ویژگی‌های جریان محلول و دور پمپ داشتند، رکوردهایی که نوع صافی آن‌ها ثبت نشده بود، رکوردهایی که عدد دور پمپ در طول دیالیز برای آن‌ها ثابت نبود و رکوردهای تکراری متعلق به یک بیمار، حذف گردیدند. بنابراین، تعداد رکوردهای مورد بررسی به ۲۱۴ مورد رسید. همچنین بعضی از اطلاعات موجود در پرونده بیماران مانند سن و جنس به دلیل عدم ثبت در اکثر پرونده‌ها و نام و نام خانوادگی، حذف شدند. میانگین‌های وزن بیماران قبل و بعد دیالیز به ترتیب ۵۸/۹ و ۵۷/۱ کیلوگرم، فشارخون سیستولیک بیمار قبل و بعد از دیالیز به ترتیب ۱۴۰ و ۱۳۷/۸ میلی‌متر جیوه و فشار خون دیاستولیک بیمار قبل و بعد از دیالیز به ترتیب ۸۴/۵ و ۸۲/۵ میلی‌متر جیوه است. مدل‌سازی با استفاده از نرم افزار کلمنتاین ۱۲ انجام شده است. روش کار در پژوهش حاضر داده کاوی پیش‌بینانه بود. در الگوریتم‌های پیش‌بینی هدف پیش‌بینی یک ویژگی خاص بر مبنای ویژگی‌های دیگر است. ویژگی پیش‌بینی شونده، یک متغیر وابسته و بقیه متغیرها مستقل نامیده می‌شوند [۱۶].

الگوریتم‌های درخت تصمیم و شبکه عصبی برای مدل‌سازی استفاده شد. در تولید مدل، ابتدا داده‌های تحت بررسی به طور تصادفی به دو مجموعه مستقل آموزش و آزمایش تقسیم گردید. داده‌های بخش آموزش (۸۰ درصد) مدل را تولید و داده‌های بخش آزمایش (۲۰ درصد) مدل تولید شده را آزمایش و برچسب مربوط به رکوردهای مذکور را تعیین نمود. برای آموزش درخت تصمیم یک متغیر رشته‌ای (طبقه ای/اسمی) فیلد خروجی بود و یک یا تعداد بیشتری فیلد ورودی وجود داشت. جدول یک صفات مورد استفاده در مدل‌سازی را نشان می‌دهد. پارامترهای وزن، فشار خون و ضربان قلب پارامترهای مورد توجه در دیالیز هستند زیرا در ابتدای هر جلسه ثبت می‌گردند [۱۷]. جریان محلول، دور پمپ و سطح غشاء نیز پارامترهایی هستند که در تعیین دیالیز خوب، نقش دارند [۱۸]. از این‌رو در پژوهش حاضر صفات وزن، فشارخون، جریان محلول و دور پمپ برای تعیین نوع صافی مناسب مورد بررسی قرار گرفت. برای اطمینان از کافی بودن صفات انتخاب شده، از نظر سه نفر از مدیران بخش دیالیز بیمارستان‌ها بهره گرفته شد.

جدول ۱: داده‌ها و نوع آن‌ها پس از پاک‌سازی

مشخصه	نوع	رده مشخصه
وزن بیمار قبل دیالیز	عددی	۱۳۰-۱۵/۸
فشار خون سیستولیک بیمار قبل دیالیز	عددی	۶۷-۲۳۰
فشارخون دیاستولیک بیمار قبل دیالیز	عددی	۱۰-۱۴۸
جریان محلول	عددی	۱۰۲-۸۰۰
دور پمپ	عددی	۱۳۰-۳۷۰
نوع صافی	اسمی	F6, F7, F8, F60, F70, F80, PS10 LF, PS13 LF, PS16 LF, PES13 LF, PS130 HF

دسته‌بندی کاربرد دارند در حالی که درخت‌های تقسیم و رگرسیون و سی ۴/۵ (C4. 5) در حل هر دو نوع مسائل رگرسیون و دسته‌بندی مورد استفاده قرار می‌گیرند [۱۹]. سی فایونسخه جدید سی ۴/۵ است که نسبت به سی ۴/۵ حافظه کمتری هنگام تولید مجموعه قوانین استفاده می‌نماید [۲۰].

برای مدل‌سازی از الگوریتم‌های درخت تصمیم سی فایو، تقسیم و رگرسیون، آشکارساز تعامل خودکار مجذور کای، درخت آماری کارا و بی‌طرف سریع و شبکه عصبی استفاده گردید. درخت تصمیم آشکارساز تعامل خودکار مجذور کای و درخت آماری کارا و بی‌طرف سریع برای حل مسائل

بنابراین، مسائلی که با دسته‌های نامتعادل سروکار دارند و یا در مسائلی که ارزش دسته‌ای در مقایسه با دسته دیگر متفاوت است، از معیارهای دیگری استفاده می‌شود. در مسائل واقعی معیارهای دیگری نظیر حساسیت و نرخ مثبت غلط اهمیت ویژه‌ای دارند. این معیارها که توجه بیشتری به دسته بندی مثبت نشان می‌دهند، توانایی دسته بند را در تشخیص دسته مثبت تبیین می‌کنند. معیار حساسیت نشان می‌دهد که دقت تشخیص دسته مثبت چه مقدار است و معیار نرخ مثبت غلط نرخ هشدار غلط را با توجه به دسته منفی بیان می‌کند [۲۲]. شاخص‌های مورد نظر طبق روابط یک تا ۱۱ محاسبه می‌گردند [۲۳].

$TP$  = تعداد داده‌های برچسب مثبتی که درست دسته بندی شده‌اند،

$FP$  = تعداد داده‌های برچسب منفی که به نادرست مثبت دسته بندی شده‌اند،

$FN$  = تعداد داده‌های برچسب مثبتی که به نادرست منفی دسته بندی شده‌اند،

$TN$  = تعداد داده‌های برچسب منفی که درست دسته بندی شده‌اند.

گزینه آزمایشی انتخاب شده برای الگوریتم‌های درخت تصمیم اعتبار سنجی متقاطع با ده تکرار (10-Fold Cross-Validation) بود زیرا آزمایش‌ها نشان داده‌اند که بهترین انتخاب برای به دست آوردن دقیق‌ترین تخمین، اعتبار سنجی متقاطع ده است [۲۱]. درصد نويز قابل انتظار برای این الگوریتم‌ها، صفر و تعداد سطوح زیر ریشه، پنج انتخاب گردید. شبکه عصبی مصنوعی نیز از یک لایه مخفی با ۲۰ نرون، حالت آموزش دسته‌ای و الگوریتم‌های سریع با پایداری ۲۰۰ بهره می‌برد. برای ارزیابی مدل‌ها و انتخاب بهترین مدل از شاخص‌های حساسیت (TPR)، شفافیت (TNR)، نرخ مثبت غلط (FPR)، نرخ منفی غلط (FNR)، دقت یا ارزش اخباری مثبت (PPV)، ارزش اخباری منفی (NPV)، نرخ کشف غلط (FDR)، صحت (ACC)، شاخص (FM) میانگین هندسی (GM) و نرخ خطا (ER) استفاده گردید (روابط - یک تا ۱۱).

در مسائل واقعی، معیار صحت دسته بندی، به هیچ عنوان معیار مناسبی برای ارزیابی کارایی الگوریتم‌های دسته بندی نمی‌باشد، به دلیل این که در رابطه صحت دسته بندی، ارزش رکوردهای دسته‌های مختلف یکسان در نظر گرفته می‌شوند.

$$TPR(\text{True Positive Rate or Sensitivity}) = \frac{TP}{TP+FN} \quad (۱)$$

$$TNR(\text{True Negative Rate or Specificity}) = \frac{TN}{FP+TN} \quad (۲)$$

$$FPR(\text{False Positive Rate or Fall - out}) = \frac{FP}{FP+TN} = 1 - TNR \quad (۳)$$

$$FNR(\text{False Negative Rate or Miss Rate}) = \frac{FN}{TP+FN} = 1 - TPR \quad (۴)$$

$$PPV(\text{Positive Predictive Value or Precision}) = \frac{TP}{TP+FP} \quad (۵)$$

$$NPV(\text{Negative Predictive Value}) = \frac{TN}{TN+FN} \quad (۶)$$

$$FDR(\text{False Discovery Rate}) = \frac{FP}{FP+TP} = 1 - PPV \quad (۷)$$

$$ACC(\text{Accuracy}) = \frac{TP+TN}{TP+FN+FP+TN} \quad (۸)$$

$$FM(F - Measure) = \frac{2 \times TP}{TP+FP+FN} \quad (۹)$$

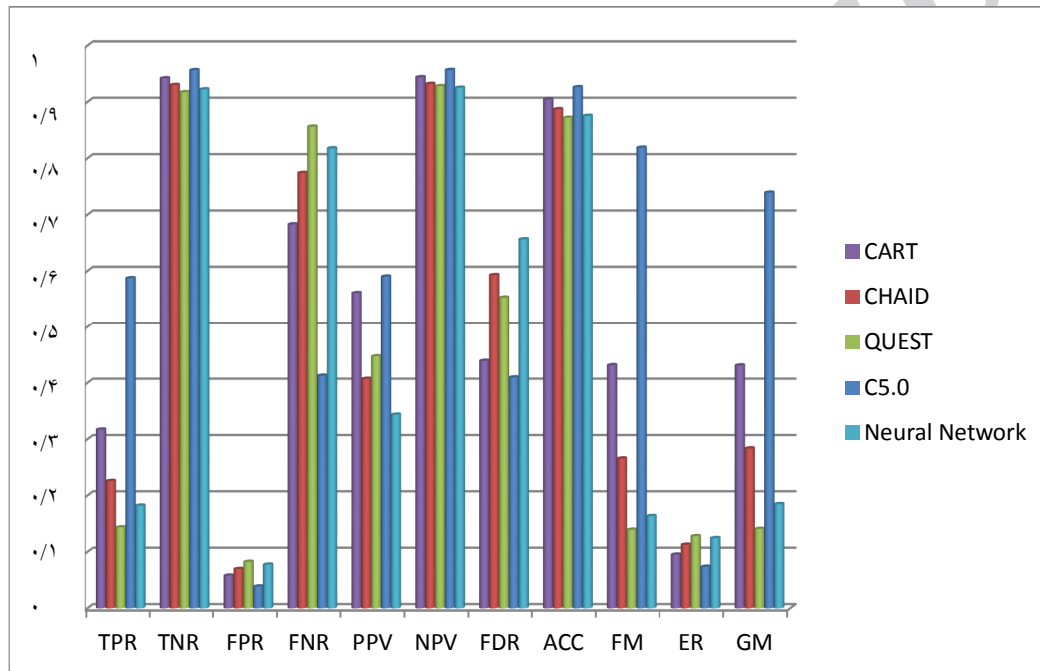
$$ER(\text{Error Rate}) = \frac{FP+FN}{TP+FN+FP+TN} = 1 - \text{Accuracy} \quad (۱۰)$$

$$GM(\text{Geometric Mean}) = \sqrt{TPR \times TNR} \quad (۱۱)$$

## یافته‌ها

شاخص F و میانگین هندسی برای این مدل دارای بیشترین مقدار است. مقادیر این شاخص‌ها هرچه بیشتر باشد نشان‌دهنده این است که طبقه بند مورد استفاده نمونه‌های بیشتری را در جای درست خود طبقه بندی کرده است. شاخص‌های نرخ-های مثبت غلط، منفی غلط، کشف غلط و خطا برای این مدل کمترین مقدار را دارند. مقادیر کم این شاخص‌ها تایید کننده وقوع خطای کمتر در طبقه بندی نمونه‌ها است. شکل یک نمودار شاخص‌ها را جهت مقایسه بهتر نشان می‌دهد.

یافته‌ها نشان می‌دهد ۱۹/۱ درصد بیماران دارای فشار طبیعی، ۳۹/۲ درصد دارای پیش هیپرتانسیون، ۱۸/۷ درصد در مرحله یک، ۱۴/۹ درصد در مرحله دو و ۷/۹ درصد دارای هیپرتانسیون سیستمولیک مجزاهستند. مقادیر شاخص‌های ارائه شده در جدول دو روی مجموعه داده‌های آزمایش نشان داد که گره سی فایو بهترین مدل را تولید نموده است. شاخص-های حساسیت، شفافیت، صحت، دقت، ارزش اخباری منفی،



شکل ۱: نمودار شاخص‌ها برای مدل‌های تولید شده

رتبه‌ای و یا کمی می‌باشد، استفاده می‌گردد [۲۴]. این معیار برای مقایسه مدل‌های دسته بندی به طور گرافیکی استفاده می‌شود و نتیجه حاصل از این تحلیل، منحنی‌هایی هستند که با استفاده از آن‌ها، به راحتی می‌توان تشخیص داد کدام مدل مناسب‌تر است. منحنی ایجاد شده مبادله بین حساسیت و شفافیت-۱ است [۲۵].

پارامترهای دور پمپ، فشار خون سیستمولیک و دیاستولیک قبل دیالیز به ترتیب با مقادیر ۰/۴۶، ۰/۲۸۹ و ۰/۲۵ به‌عنوان

از بین الگوریتم‌های مورد استفاده بهترین نتایج مربوط به الگوریتم سی فایو با دقت ۰/۵۹۰۲، صحت ۰/۹۲۶۳، شفافیت ۰/۹۵۶۶ و حساسیت ۰/۵۸۷۲ است. در تحلیل داده‌های پزشکی، استفاده از معیارهای شفافیت و حساسیت برای ارزیابی مدل پیش‌بینی کننده مناسب است [۳]. همچنین جهت ارزیابی کارایی دسته بندها از روش موثر و شناخته شده منحنی مشخصه عملکرد سیستم (Receiver Operating Characteristic: ROC) که نتایج آن متغیری در مقیاس

استخراج شده می‌تواند به‌عنوان یک سیستم تصمیم‌یار در دنیای واقعی مورد استفاده قرار گیرد [۱۲].

یافته‌های این پژوهش نشان داد از بین الگوریتم‌های مورد استفاده، بهترین نتایج از درخت سی فایو با دقت ۰/۵۹۰۲، صحت ۰/۹۲۶۳، شفافیت ۰/۹۵۶۶ و حساسیت ۰/۵۸۷۲ جهت پیش‌بینی نوع صافی همودیالیز حاصل گردید. سپهری و همکاران از درخت تصمیم برای ساخت مدلی جهت تعیین نوع درمان سنگ حالب بیماران مبتلا به سنگ کلیه، استفاده نمودند. صحت مدل تولید شده حدود ۸۵ درصد و با میانگین -گیری صحت K مدل تولید شده، به‌دست آمد در حالی که الگوریتم درختی مورد استفاده در بیمارستان عملکرد ۷۷ درصد داشت [۱۲]. یای و همکاران از الگوریتم‌های اپریوری و درخت تصمیم سی فایو برای کشف الگوی بستری شدن بیماران همودیالیزی استفاده نمودند. نتایج تجربی نشان داد که مقدار حساسیت برای درخت تصمیم ایجاد شده روی داده‌های شش ماه بیماران ۸۱/۶۰ درصد و روی داده‌های سه ماه ۱۳ درصد است. کمترین مقدار پشتیبان برای الگوریتم اپریوری نیز یک دهم ۰ در نظر گرفته شد [۳].

دانش کشف شده به فرم قوانین تصمیم‌گیری است و متغیر وابسته نوع صافی بر اساس متغیرهای مستقل وزن قبل دیالیز، فشار خون دیاستولیک و سیستولیک قبل دیالیز، جریان محلول و دور پمپ بیان شده است. درصد صحت هر قانون، درصد نمونه‌هایی است که توسط قانون مورد پذیرش قرار گرفته‌اند و در درخت تصمیم در جای درست خود قرار گرفته‌اند. درصد صحت بالای ۵۰ درصد برای ۹۵ درصد قوانین موید میزان پیش‌بینی درست مدل تولید شده است [۹]. کوزیاک و همکاران دانش کشف شده به فرم قوانین تصمیم‌گیری را از دو الگوریتم درخت تصمیم و مجموعه خشن استخراج نمودند. الگوریتم درخت تصمیم ۶۷ درصد و الگوریتم مجموعه خشن ۵۶ درصد پیش‌بینی درست روی مجموعه آزمایشی داشت [۱۵]. بلازی و همکاران دو روش برای استخراج قواعد انجمنی بالقوه بین الگوهای چند متغیره ارائه نمودند که رفتار و مشکلات بیماران را مشخص می‌کرد. این

تأثیرگذارترین پارامترها در مدل‌سازی توسط درخت تصمیم سی فایو انتخاب گردیدند. قوانین تولید شده نیز پارامترهای تأثیرگذار روی نوع صافی دیالیز را شناسایی و تأیید نمود و توسط متخصصان حوزه دیالیز مورد بررسی و اعتبار آن‌ها با در نظر داشتن مقدار درصد صحت قانون، مورد تأیید قرار گرفت. جدول سه قوانین استخراج شده از درخت تصمیم سی فایو را نشان می‌دهد. قوانین ایجاد شده برای یک نمونه جدید با ویژگی‌های مشخص، می‌توانند صافی مناسب برای دیالیز را پیش‌بینی نمایند. از آنجایی که فشار خون دومین عامل بروز بیماری‌های کلیوی است، اطلاعات زیر از قوانین ایجاد شده، استخراج گردید:

- ۱- صافی مناسب برای بیماران دارای فشار خون طبیعی F60, F70, PS13, PS16 است.
- ۲- صافی مناسب برای بیماران دارای پیش‌هیپرتانسیون F6, F7, F60, F70, F80, PES13, PS13 و PS16 است.
- ۳- صافی مناسب برای بیماران دارای فشار خون مرحله F6, F7, F60, PS13, PS130 و PS16 است.
- ۴- صافی مناسب برای بیماران دارای فشار خون مرحله F60, F70, PES13, PS13, PS130 و PS16 است.
- ۵- صافی مناسب برای بیماران دارای هیپرتانسیون سیستمیک مجزا F60, F7, PES13, PS13 و PS130 است.

### بحث و نتیجه‌گیری

یکی از شایع‌ترین راه‌های درمان بیماران دچار نارسایی کلیوی، همودیالیز است. تجویز اشتباه یا نامناسب می‌تواند به عوارض شدید در بیماران و در نتیجه قصور پزشکی منجر شود. داده‌کاوی یکی از ابزارهایی است که به پزشکان در جهت تصمیم‌گیری و سازمان‌دهی داده‌ها کمک می‌نماید [۱۶] و هدفش دستیابی به نتایجی است که از آن‌ها در دنیای واقعی برای بهبود کارایی سازمان‌ها استفاده گردد. زیرا دانش

قواعد بالقوه در برنامه درمانی موجود از قبل، استخراج و اجرا نشده بودند [۸].

متغیرهای دور پمپ، فشار خون سیستولیک و دیاستولیک قبل دیالیز در مدل ایجاد شده به عنوان پارامترهای تاثیرگذار شناخته شدند. گنس ولر و همکاران نشان دادند که دور پمپ بر میزان کفایت دیالیز و برداشت پتاسیم موثر است [۲۶]. در الگوریتم ارائه شده توسط سپهری و همکاران سن بیمار در تصمیم‌گیری موثر بود در حالی که در الگوریتم بیمارستان برای انتخاب نوع درمان، دو ویژگی اندازه و مکان سنگ مهم بود [۱۲].

در برخی بیمارستان‌ها الگوریتم‌های درخت مانندی وجود دارد که تصمیم‌گیری در برخی زمینه‌ها براساس آن انجام می‌شود، بنابراین، ساخت مدلی که خروجی آن به صورت یک درخت نمایش داده شود می‌تواند به خوبی در بیمارستان مورد استفاده قرار گیرد [۱۲]. در مطالعه‌ی حاضر خروجی مدل سی فایو به صورت یک درخت تصمیم‌گیری است که قوانین استخراج شده از آن راهنمای مناسبی برای مدیریت درمان بیماران همودیالیز است. چیدرو و همکاران مجموعه‌ای از ابزارهای تعاملی برای بصری‌سازی سیستم‌های همودیالیز سنتی توسعه دادند. آن‌ها رویکردی برای داده‌کاوی بصری روی داده‌های زمانی ابداع نمودند و از آن برای حل مشکلات واقعی پزشکی از جمله مدیریت بیماران همودیالیز استفاده نمودند [۱۴].

اهمیت نفوذپذیری صافی دیالیز در کفایت دیالیز مساله‌ای انکارناپذیر است. عشوندی و همکاران به بررسی کفایت دیالیز با صافی‌های فلاکس در بیماران همودیالیزی پرداختند. نتایج این پژوهش نشان داد که کفایت دیالیز واحدهای پژوهش با استفاده از صافی‌های فلاکس در ۵۰ درصد موارد کاملاً مطلوب و فقط در ۱۰ درصد موارد ناکافی بوده است [۲]. مسلم و همکاران نیز به بررسی کفایت دیالیز و ارتباط آن با نوع صافی مورد استفاده پرداختند. در این پژوهش هر دو نوع صافی‌های فلاکس و لوفلاکس به صورت ضربدری و به مدت یک‌ماه برای بیماران استفاده گردید. نتایج

نشان داد که ۶۰ درصد بیمارانی که از صافی‌های لوفلاکس و ۸۰ درصد بیمارانی که از صافی‌های‌های فلاکس استفاده نمودند، کفایت دیالیز بیش از ۱/۲ داشتند. این مطالعه نشان داد که ارتباط آماری معناداری بین میزان هموگلوبین و نوع صافی وجود دارد [۱]. واحدپرست و روانی پور استفاده از صافی‌های لوفلاکس را یکی از عوامل پایین بودن کفایت دیالیز برشمردند [۶].

انتخاب درست نوع صافی بر تعداد دفعات یا مدت زمان دیالیز تاثیرگذار است. برزو و همکاران توجه به استفاده از صافی‌های متناسب با وزن بیمار را لازم دانستند [۷]. عابدی سماکوش و همکاران پیشنهاد دادند به منظور ارتقای کفایت دیالیز و بهبود زندگی بیماران دیالیزی، دوز دیالیز برای هر فرد به طور اختصاصی محاسبه شود و مدت زمان تجویز دیالیز، بر اساس نیاز بیمار تنظیم گردد [۵]. کچال و همکاران نشان دادند که انتخاب صافی دیالیز بر گردش سطح بتا-۲-میکروگلوبین مستقل از زیست‌سازگاری و تراوایی غشای دیالیزور مورد استفاده، تاثیرگذار است [۲۷]. طاوسی و همکاران نشان دادند که مقدر داروی تجویزی اریتروپویتین برای بیماران همودیالیزی در هر ماه بر مقدار دارو در ماه بعدی تاثیر مستقیم دارد [۱۶].

بررسی مطالعات خارجی و داخلی نشان می‌دهد که تاکنون از داده‌کاوی برای آنالیز داده‌های دیالیز استفاده گردیده است. نتایج حاصل شده کاربردی بودن داده‌کاوی در حوزه دیالیز را تایید می‌نماید. از نظر نوع تکنیک مورد استفاده و با توجه به بکارگیری بیش‌تر تکنیک‌های دسته‌بندی در مطالعات پیشین و نیز تفسیرپذیری و قابل فهم بودن نتایج حاصل از آن‌ها، این تکنیک‌ها به عنوان روش پیش‌بینی انتخاب شد. سپس با توجه به متغیر هدف (نوع صافی)، الگوریتم‌هایی که قابل استفاده بود روی داده‌ها آزمون شد تا بهترین الگوریتم شناسایی شود. یافته‌ها نشان می‌دهد که درخت سی فایو در شاخص‌های حساسیت، شفافیت، صحت، دقت، ارزش اخباری منفی، شاخص F و میانگین هندسی بیشترین مقدار را دارد و نسبت به سایر الگوریتم‌ها تعداد بیشتری از نمونه‌ها را در جای درست



تأمین تجهیزات بیمارستانی می‌باشد که با انتخاب سطح موجودی مناسب می‌توان احتمال مواجهه با کمبود تجهیزات در مراکز درمانی را کاهش و برنامه‌ریزی سیستماتیک برای سازماندهی تجهیزات مورد نیاز و پرمصرف به شیوه‌ای موثر انجام داد. از آنجایی که نتایج این تحقیق وابسته به داده‌های یک بیمارستان می‌باشد، پیشنهاد می‌شود برای بررسی بیشتر در این زمینه، در مطالعات بعدی از داده‌های مراکز درمانی دیگر و سایر الگوریتم‌ها استفاده و نتایج با هم مقایسه گردد تا میزان ارتقای عملکرد سازمان‌های سلامت بهتر مورد سنجش قرار گیرد. عوامل سن و جنس می‌تواند به عنوان عناصر تاثیرگذار بر درمان مورد بررسی قرار گیرند و بهتر است در مطالعات بعدی این عوامل در نظر گرفته شود.

### تشکر و قدردانی

این مقاله حاصل طرح تحقیقاتی مستقل و بدون حمایت سازمانی است.

خود دسته بندی نموده است. از آنجایی که عملکرد این مدل می‌تواند بهتر از مدل ذهنی پزشکان باشد، از آن می‌توان به - عنوان سیستمی تصمیم‌یار جهت تعیین نوع صافی همودیالیز در بیمارستان استفاده کرد.

از نظر هدف در مطالعات خارجی و داخلی، پیش‌بینی نوع صافی همودیالیز کمتر مورد توجه قرار گرفته است. با توجه به نتایج پژوهش، در مراکز همودیالیز؛ وجود رویکردی سازمان-دهی شده جهت پیش‌بینی نوع صافی همودیالیز به‌منظور کمک به پرسنل بخش دیالیز در جهت افزایش صحت تشخیص و جلوگیری از عوارض جانبی تشخیص نادرست، ضروری است. نقطه قابل بهبود در این حیطة پیش‌بینی دقیق صافی مورد نیاز برای همودیالیز به جهت حفظ رفاه حال بیماران با افزایش کفایت و کارآیی همودیالیز و کاهش مرگ و میر این بیماران و ایجاد سیستم‌های تصمیم‌یار جهت کمک در تشخیص اولیه نوع صافی می‌باشد. از محدودیت‌های مطالعه حاضر می‌توان به عدم وجود نمونه برای برخی از صافی‌ها در مجموعه داده‌ی تحت بررسی اشاره نمود. عدم وجود نمونه برخی صافی‌ها سبب می‌گردد در مدل تولید شده جایی برای آن نمونه‌ها در نظر گرفته نشود و قانونی برای آن صافی‌ها تولید نگردد.

در این مقاله درخت تصمیم سی فایو تولید شده دارای بیشترین مقدار شاخص‌های حساسیت، شفافیت، صحت، دقت، ارزش اخباری منفی، شاخص F و میانگین هندسی و کمترین مقدار شاخص‌های نرخ مثبت غلط، نرخ منفی غلط، نرخ کشف غلط و نرخ خطا است و برای پیش‌بینی مناسب‌تر می‌باشد. متغیرهای دور پمپ، فشار خون سیستمولیک و دیاستولیک قبل دیالیز به عنوان پارامترهای تاثیرگذار شناخته شدند.

پژوهش حاضر از روش‌های پیش‌بینی برای استخراج قوانین مرتبط با سیستم‌های خبره استفاده نموده است. سیستم‌های خبره علاوه بر کمک به پرسنل بخش دیالیز جهت پیش‌گیری از انتخاب نادرست صافی، می‌توانند در پیش‌بینی تعداد صافی مورد نیاز مراکز درمانی با توجه به شمار بیمارانی مفید واقع شود که به‌طور عمده به این مراکز مراجعه می‌کنند. عدم قطعیت در پیش‌بینی تقاضای تعداد صافی یکی از مشکلات

## References

1. Moslem A, Naghavi M, Basiri Moghadam M, Gharche M, Basiri Moghadam K. [Assessing the adequacy of dialysis and its relationship with kind of filter in patient under hemodialysis referred to 22-Bahman hospital of Gonabad]. *Horizon Med Sci* 2008; 14(2): 20-23. [Persian]
2. Oshvandi K, Kavannejad R, Borzuo SR, Gholyaf M, Salavati M. [Dialysis Adequacy with High Flux Membrane in Hemodialysis Patients at Shahid Beheshti Hospital, Hamedan]. *J Urmia Nurs Midwifery Fac* 2012; 10(4): 540-548. [Persian]
3. Yeh JY, Wu TH, Tsao CW. Using data mining techniques to predict hospitalization of hemodialysis patients. *Decision Support Systems* 2011; 50(2): 439-448.
4. Saeedi M, Zareie F, Javaheri F. [The assessment of Dialysis adequacy and its related factors among hemodialysis patients]. *Aligoodarz Nursing Faculty Analytic Research Journal* 2015; 6(3-4): 23-31. [Persian]
5. Abedi Samakoosh M, Aghaie N, Gholami F, Shirzad M, Yosefi E, Teymoorzadeh Baboli M. [Assesment Dialysis Adequacy in Hemodialysis Patients of Qaemshar Razi Hospital in 2012]. *J Mazand Univ Med Sci* 2013; 23(107): 20-27. [Persian]
6. Vahed Parast H, Ravanipour M. [Assessing the Adequacy of Dialysis in Patients Undergoing Hemodialysis in Hemodialysis Center in Boshehr City]. *Scientific Journal of Hamadan Nursing & Midwifery Faculty* 2008; 16(2): 50-60. [Persian]
7. Borzu SR, Ghalyaf M, Amini R, Zandeye M, Torkaman B. [The effect of increasing blood flow rate on dialysis adequacy in hemodialysis patients]. *J Shahrekord Univ Med Sci* 2006; 8(2): 60-66. [Persian]
8. Bellazzi R, Larizza C, Magni P, Bellazzi R. Temporal data mining for the quality assessment of hemodialysis services. *Artificial Intelligence in Medicine* 2005; 34(1): 25-39.
9. Jooriyan N, Ashoori M. [Predicting the Effectiveness of Preeclampsia Medications Based on Dose and Method of drug Consumption Using Data Mining]. *Iranian Journal of Obstetrics, Gynecology and Infertility* 2014; 17(123): 13-22. [Persian]
10. Ashoori M, Alizade S, Hosseiny Eivary HS, Rastad S, Hosseiny Eivary SS. A model to predict the sequential behavior of healthy blood donors using data mining. *Journal of Research & Health* 2015; 5(2): 141-148.
11. Ashoori M, NajiMoghadam V, Alizadeh S, Safi M. [Classification and Clustering Algorithm Application for Prediction of Tablet Numbers: Case Study Diabetes Disease]. *Health Inforamtion Management* 2013; 10(5): 739-749. [Persian]
12. Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. [A data mining based model for selecting type of treatment]. *Tehran Univ Med J* 2009; 67(6): 421-427. [Persian]
13. Fernandez EA, Valtuille R, Willshaw P, Perazzo CA. Dialysate-side urea kinetics. Neural network predicts dialysis dose during dialysis.

- Medical & Biological Engineering & Computing 2003; 41(4): 392-396.
14. Chittaro L, Combi C, Trapasso G. Data mining on temporal data: a visual approach and its clinical application to hemodialysis. Journal of Visual Languages and Computing 2003; 14(6): 591-620.
15. Kusiak A, Dixon B, Shah S. Predicting survival time for kidney dialysis patients: a data mining approach. Computers in Biology and Medicine 2005; 35(4): 311-327.
16. Tavousi A, Sepehri MM, Malakoutian T, Khatibi T. [Data Mining Approach in Prediction of Erythropoietin Dosage in Hemodialysis Patients]. J Mazandaran Univ Med Sci 2015; 25(129): 26-35. [Persian]
17. Munger MA, Ateshkadi A, Cheung AK, Flaharty KK, Stoddard GJ, Marshall EH. Cardiopulmonary Events During Hemodialysis: Effects of Dialysis Membranes and Dialysate Buffers. American Journal of Kidney Diseases 2000; 36(1): 130-139.
18. Ahmed HM, Aquina CT, Gracias VH. Damage Control Surgery. In GIBNEY RTN. Robert Wood Johnson University Hospital. New Brunswick: Springer; 2012. 712-713.
19. Pham H, editor. Springer Handbook of Engineering Statistics. Germany: Springer; 2006.
20. Pandya R, Pandya J. International Journal of Computer Applications. C5. 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning 2015; 117(16): 18-21.
21. Ameri H, Alizade S, Barzegari A. [Knowledge Extraction of Diabetics' Data by Decision Tree Method]. Journal of Health Administration 2013; 16(53): 58-72. [Persian]
22. Seliya N, Khoshgoftaar T. The use of decision trees for cost-sensitive classification: an empirical study in software quality prediction. WIRES Data Mining and Knowledge Discovery 2011; 1(5): 448-459.
23. Han J, Kamber M, Pai J, editors. Data Mining: Concepts and Techniques. 3rd ed. USA: Morgan Kaufman; 2000.
24. Sadat-Hashemi S, Ghorbani R, Kavehie B. [Analyzing receiver operating characteristic curves to compare medical diagnostic tests]. koomesh 2005; 6(2): 145-150. [Persian]
25. Lowry R. Simple ROC Curve Analysis. VassarStats: Website for Statistical Computation 1998-2016 [cited 2001-2016]. Available From: URL: <http://vassarstats.net/roc1.html>.
26. Gutzwiller J, Schneditz D, Huber A, Schindler C, Garbani E, Zehnder C. Increasing blood flow increase kt/v and potassium removal but fails to improve phosphate removal. Clin Nephrol 2003; 59(2): 130-136.
27. Küchle C, Lang SM, Schiffli H. Effect of Dialysate Buffer on Serum Beta-2-Microglobulin Levels in Chronic Hemodialysis. Nephron 1998; 79(1): 99-100.



# A Model to Predict Hemodialysis Buffer Type Using Data Mining Techniques

Ashoori M<sup>1</sup>

## Abstract

**Introduction:** Inadequate dialysis for patients' kidneys as a mortality risk necessitates the presence of a pattern to assist staff in dialysate part to provide the proper services for dialysis patients and also the proper management of their treatment. Since the role of buffer type in the adequacy of dialysis is determinative, the present study is aimed at determining hemodialysis buffer type.

**Methods:** Cross-sectional method was applied in the present study. The population included the data extracted from Ali Ibn Abi Talib hospitals in Zahedan from May-June 2016. In this study Clementine 12.0 has been used for data analysis. In the present study Decision trees C5.0, Classification and Regression Tree, Chi-Squared Automatic Interaction Detector, Unbiased and Efficient Statistical Tree and Neural Network algorithms were executed.

**Results:** The obtained accuracy for executing decision trees C5.0, Classification and Regression Tree, Chi-Squared Automatic Interaction Detector, Unbiased and Efficient Statistical Tree and Neural Network equals 0.9263, 0.9047, 0.8872, 0.8720 and 0.8754, respectively. The results of indices including sensitivity, specificity, accuracy, precision, NPV, FM, GM, FPR, FNR, FDR, ER for C5.0 decision tree are indicators of better performance of this algorithm compared to the other algorithms.

**Conclusion:** The extracted rules for a new sample having specific features can predict proper dialysis buffer. Hence, the proposed model helps us in predicting more precise hemodialysis buffer type and also the proper management of patient treatment which result in better performance among health organization.

**Keywords:** Hemodialysis Buffers, Data Mining, Decision Tree, Neural Network

• Received: 21/July/2016 • Modified: 8/Nov/2016 • Accepted: 5/Feb/2017

1. Instructor of Information Technology Department, School of Technical and Engineering, Higher Educational Complex of Saravan, Saravan, Iran; Corresponding Author (mashoori@saravan.ac.ir)