ارایه مدلی جهت پیش‌بینی بقای بیماران دیالیز صفاقی با استفاده از الگوریتم‌های داده‌کاوی

فرزاد فیروزی جهانتیغ[1*]
ایرج نجفی[2]
مریم استواره[1]

۱- گروه مهندسی صنایع، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان، ایران.
۲- گروه نفرولوژی، بیمارستان شریعتی، دانشگاه علوم پزشکی تهران، تهران، ایران.

[*] نویسنده مسئول: سیستان و بلوچستان، زاهدان، بلوار دانشگاه، دانشگاه سیستان و بلوچستان، دانشکده مهندسی، گروه مهندسی صنایع.
تلفن: ۲-۸۸۹۰۵۰۱۱-۰۲۱
E-mail: Firouzi@eng.usb.ac.ir

چکیده

دریافت: ۱۳۹۶/۰۶/۰۸ ویرایش: ۱۳۹۶/۱۰/۰۹ پذیرش: ۱۳۹۶/۱۰/۱۴ آنلاین: ۱۳۹۶/۱۰/۱۵

زمینه و هدف: دیالیز صفاقی یکی از متداول‌ترین روش‌های درمان بیماران مبتلابه مرحله نهایی نارسایی کلیوی می‌باشد. در سال‌های اخیر، نرخ مرگ‌ومیر بیماران تحت این درمان کاهش یافته، اما بقای طولانی‌مدت، همچنان چالش مهمی برای سیستم‌های درمانی به شمار می‌رود. پژوهش حاضر با هدف پیش‌بینی بقای بیماران دیالیز صفاقی سرپایی مداوم صورت گرفت.

روش بررسی: در این مطالعه با توجه به تفاوت اهمیت نسبی فاکتورها در بیماران مختلف، ابتدا عوامل مؤثر در بقای بیماران دیالیز صفاقی توسط الگوریتم تصادفی فارست (Random forests analysis) شناسایی شده‌اند. سپس با استفاده از الگوریتم‌های چند کلاسه ماشین بردار پشتیبان یکی در برابر همه و ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چندفضایی درخت باینری، داده‌های بالینی و آزمایشگاهی بیماران تحت درمان دیالیز صفاقی سرپایی مداوم به‌صورت گذشته‌نگر از مرداد ماه ۱۳۷۵ تا فروردین ماه ۱۳۹۳ در ۱۸ مرکز دیالیز صفاقی ارزیابی‌شد.

یافته‌ها: تعداد ۳۰۹۷ بیمار با میانگین سنی ۱۵/۶۷±۵۰/۶۳ سال و متوسط زمان پیگیری ۱۹/۱۳±۲۴/۴۸ ماه، مورد مطالعه قرار گرفتند. نتایج حاصل از الگوریتم رندوم فارست ۳۵ عامل را به‌عنوان مهمترین عوامل پیش‌بینی کننده بقای تعیین نموده است. سپس پیش‌بینی وضعیت بقای بیماران دیالیز صفاقی با استفاده از دو الگوریتم طبقه‌بندی چند کلاسه ماشین بردار پشتیبان، در پنج کلاس بیماران "باقی‌مانده در روش دیالیز صفاقی"، "انتقال به همودیالیز"، "دریافت پیوند کلیه"، "فوت شده" و "بهبود عملکرد کلیه" ارزیابی‌شد. درستی الگوریتم‌های پیش‌بینی به‌ترتیب ۵۱/۹۹٪ و ۸۹/۵۷٪ به‌دست آمد.

نتیجه‌گیری: الگوریتم ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چندفضایی درخت باینری، با در نظر گرفتن شاخص‌های ارزیابی متعدد و توابع توزیع متفاوت کلاس‌ها، از دقت بالایی جهت پیش‌بینی بقای بیماران دیالیز صفاقی سرپایی مداوم برخورداراست.

کلمات کلیدی: دیالیز صفاقی سرپایی مداوم، داده‌کاوی، وضعیت سلامت، پژوهش‌های گذشته‌نگر، میزان بقاء.

مقدمه

امروزه تعداد بیماران مرحله نهایی بیماری کلیوی (End stage renal disease, ESRD) با توجه به جمعیت سالخورده، امید به زندگی طولانی‌تر بیماران، افزایش دسترسی به درمان‌های جایگزین کلیوی و بروز بیماری‌هایی چون دیابت و فشارخون، رو به افزایش می‌باشد.[1]

دیالیز صفاقی (Peritoneal dialysis, PD)، یکی از روش‌های درمانی جذاب و اقتصادی برای درمان نارسایی کلیه است که در اواخر ۱۹۷۰ تشریح گردید.[2] بقا و کیفیت زندگی گزارش شده از بیماران تحت این درمان، مشابه بیمارانی است که از درمان همودیالیز (Hemodialysis, HD) استفاده می‌کنند.[3و4] به‌علاوه ماهیت خانگی و

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

انعطاف‌پذیری دیالیز صفاقی، مزیت استفاده از این روش نسبت به همودیالیز در بسیاری از مراکز سراسر جهان می‌باشد.[۵]

در سال‌های اخیر، نرخ مرگ‌ومیر بیماران تحت درمان دیالیز صفاقی سرپایی مداوم (Continuous ambulatory peritoneal dialysis, CAPD)، کاهش‌یافته است، اما میزان بقای طولانی‌مدت آن‌ها همچنان مسئله‌ای نگران‌کننده است، به‌طوری‌که بقای کلی بیماران در سال اول دیالیز صفاقی ۸۶/۸٪ و در بیمارانی که بیش از ۱۰ سال زنده می‌مانند، حدود ۱۱/۳٪ می‌باشد.[۶] این کاهش چشمگیر موجب تبدیل بقای بیماران دیالیز صفاقی به یکی از اولویت‌های اصلی پژوهش‌های علوم پزشکی شده است و حجم زیادی از داده‌ها را در بردارد که به‌منظور استفاده مؤثر از این داده‌ها باید روش‌های دقیق و کارآمد به‌کار گرفته شود.

داده‌کاوی، تکنیکی قدرتمند در زمینه پژوهش‌های علمی محسوب می‌شود که کامپیوترها را قادر به یادگیری داده‌ها می‌کند. استفاده از این روش، به‌منظور به‌کارگیری حجم زیادی از اطلاعات، یک رویکرد کلیدی به شمار می‌رود. به‌طوری‌که الگوریتم‌های داده‌کاوی، مؤثرترین روش‌های آنالیز مجموعه داده‌های بزرگ می‌باشند که کاربردهای متفاوتی دارند. این روش به حل مسائلی که ابزارهای آماری سنتی در پاسخ به آن ناموفق بوده‌اند، کمک می‌کند.[۷] یکی از کاربردهای این الگوریتم‌ها، شناسایی عوامل مؤثر از میان تعداد زیادی از فاکتورها است.

الگوریتم رندوم فارست (Random forests analysis) یکی از ابزارهای موفق داده‌کاوی است که به‌خوبی می‌تواند به شناسایی فاکتورهای مهم بپردازد. مهمترین ویژگی این الگوریتم عملکرد بالای آن در اندازه‌گیری اهمیت متغیرها برای تعیین نقش هر متغیر می‌باشد.[۸] کاربرد مهم دیگر این الگوریتم‌ها، توانایی بالای آن‌ها در طبقه‌بندی داده‌ها است. ماشین بردار پشتیبان، یک روش ماشین یادگیری شناخته‌شده در بسیاری از زمینه‌های طبقه‌بندی می‌باشد. اساس این روش برای طبقه‌بندی مسائل دو کلاسه، طراحی شده است، درحالی‌که مسائل دنیای واقعی اغلب نیازمند طبقه‌بندی‌های بیش از دو کلاس می‌باشند.[۹] برای گسترش ماشین بردار پشتیبان به مسائل چند کلاسه، مدل‌های طبقه‌بندی متنوعی ارایه شده که با توجه به نوع تقسیم‌بندی‌های مجموعه کل به مجموعه‌های کوچکتر، می‌تواند دقت پیش‌بینی متفاوتی را ارایه بدهد.

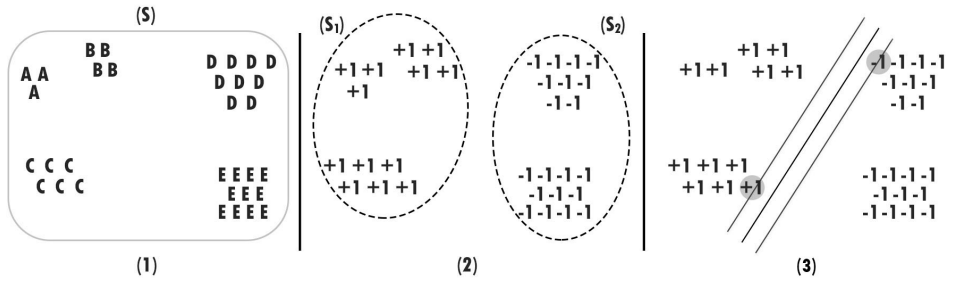
بنابراین، هدف از این مطالعه، ارزیابی بقا و پیش‌بینی وضعیت بعدی بیماران تحت درمان دیالیز صفاقی سرپایی مداوم می‌باشد. تحلیل بقا با استفاده از داده‌های بزرگ و چندمرکزی و با به‌کارگیری ماشین بردار پشتیبان چند کلاسه به‌منظور افزایش کارایی طبقه‌بندی صورت گرفته است.

روش بررسی

این پژوهش یک مطالعه هم‌گروهی گذشته‌نگر، از مردادماه ۱۳۷۵ تا فروردین ۱۳۹۳ در ۱۸ مرکز دیالیز صفاقی می‌باشد. در میان داده‌های مورد بررسی، برخی مشاهدات دارای مقادیر گمشده (Missing data)، داده‌های نویز (Noise) و داده‌های پرت (Outlier) بوده‌اند. چگونگی برخورد با این مشاهدات، به‌دلیل تاثیر نتایج حاصل از آن در تصمیم‌گیری، از اهمیت به‌سزایی برخوردار می‌باشد.[۱۰] بنابراین به‌منظور بهبود کیفیت داده‌های واقعی، ابتدا پیش‌پردازش داده‌ها صورت گرفته است. سپس از آن‌جایی که عوامل متعددی می‌تواند در بقای بیماران دیالیز صفاقی سرپایی مداوم تأثیرگذار باشد، با استفاده از الگوریتم رندوم فارست، فاکتورهای مهم شناسایی می‌گردد. اساس این روش بر مبنای ساخت چندین درخت تصمیم‌گیری به‌صورت تصادفی است.[۸] فاکتورها شامل داده‌های بالینی و آزمایشگاهی بیمارانی با مدت درمان بیش از سه ماه و محدوده سنی ۱۸ تا ۸۵ سال می‌باشند.

در نهایت، با استفاده از ماشین بردار پشتیبان چندکلاسه، ارزیابی بقای و پیش‌بینی وضعیت بعدی بیماران دیالیز صفاقی سرپایی مداوم صورت گرفته است. ابتدا روش ماشین بردار پشتیبان یکی در برابر همه (One-against-all-SVM, OAA-SVM) پیاده‌سازی گردید. روش یکی در برابر همه، یکی از مرسوم‌ترین روش‌های طبقه‌بندی چندکلاسه ماشین بردار پشتیبان می‌باشد که برای طبقه‌بندی C کلاس، به تعداد C کلاس‌بندی کننده دوتایی ماشین بردار پشتیبان احتیاج دارد. هر کلاس‌بندی کننده برای متمایز کردن یک کلاس از دیگر کلاس‌های باقی‌مانده، آموزش داده می‌شود. با این روش mامین طبقه‌بندی کننده باینری، با استفاده از داده‌های کلاس mام به‌عنوان نمونه‌های مثبت، و مابقی C-1 کلاس باقی مانده، به‌عنوان نمونه‌های منفی آموزش داده می‌شود.[۹] مشکل اصلی این رویکرد، ایجاد مجموعه

[Downloaded from tumj.tums.ac.ir on 2025-04-04]



شکل ۱: نحوه جداسازی یک نمونه پنج کلاسه به دو کلاس

آموزشی نامتوازن جهت ساخت مرز تصمیم‌گیری دقیق و تعیین برچسب کلاس بر اساس بیشینه‌ی مقدار خروجی بوده و باعث از بین رفتن تقارن مسئله اصلی می‌گردد.[۱۱] بنابراین به‌منظور بهبود بیشتر عملکرد تصمیم روش یکی در برابر همه، از ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری (Multi-space mapped binary tree-SVM, MBT-SVM) برای طبقه‌بندی چندکلاسه استفاده شده است. ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری یکی از مدل‌های طبقه‌بندی چندکلاسه می‌باشد که اساس آن، ماشین بردار پشتیبان چندکلاسه یکی در برابر همه است.[۱۲]

$$S=\{class_1, class_2, ..., class_m\} \quad , \quad m=1,2,...,C \qquad -۱$$

$$S_i = \frac{1}{l_i}\sum_{j=1}^{l_i}\sum_{k=1}^{l_i} \| x_{ij} - x_{ik} \|^2 \quad , \quad i=1,2 \qquad -۲$$

$$PF(l_1, l_2) = \frac{dis(c_1, c_2)}{S_1 + S_2} \qquad -۳$$

در این الگوریتم چگونگی تقسیم جریان نمونه‌های آموزشی به دو زیرمجموعه مجزا، بسیار اهمیت دارد. این کار با استفاده از تابع تقسیم‌بندی براساس پراکندگی داخل مجموعه صورت می‌گیرد. در این تابع، ابتدا بر اساس رابطه (۱) تمامی کلاس‌ها در مجموعه ذخیره می‌شود. سپس در رابطه (۲) تمامی کلاس‌ها با توجه به پراکندگی درون مجموعه و اندازه‌گیری نسبت توزیع داده‌های کلاس در فضای ویژگی به دو مجموعه و منتقل شده و تمامی کلاس‌های با برچسب و کلاس‌های مجموعه با برچسب در نظر گرفته می‌شود. در نهایت با رابطه (۳) یک ابرصفحه برای جدا کردن دو مجموعه و ساخته می‌شود. سپس الگوریتم ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری بر اساس تابع تقسیم بهینه، با استفاده از استراتژی اعتبار سنجی متقابل k-fold ارایه می‌شود.[۱۲]

در رابطه (۳) l_1 و l_2 به ترتیب نشان‌دهنده زیرمجموعه شامل یک کلاس و زیرمجموعه شامل تمامی کلاس‌های دیگر می‌باشد. c_i مرکز زیرمجموعه l_i و $dis(c_1, c_2)$، فاصله اقلیدسی بین c_1 و c_2 و S_i پراکندگی درون مجموعه می‌باشد. $PF(l_1, l_2)$ بزرگتر، عملکرد بهتری را در تقسیم‌بندی دارد.[۱۲] به‌منظور بهتر نشان دادن ایده تابع تقسیم‌بندی این الگوریتم، در شکل ۱، مثالی پنج کلاسه در سه بخش، مطابق با روابط (۱)، (۲) و (۳)، آورده شده است.

در این پژوهش تمامی محاسبات و تجزیه و تحلیل داده‌ها با استفاده از R Project for Statistical Computing, Version R 3.2.2 (www.r-project.org) صورت گرفته است.

یافته‌ها

در این مطالعه، تعداد ۳۰۹۷ بیمار، با میانگین سنی ۱۵/۶۷±۵۰/۶۳ سال و متوسط طول درمان ۱۹/۱۳±۲٤/٤۸ ماه، مورد مطالعه قرار گرفتند. پیش از به‌کارگیری الگوریتم‌های داده‌کاوی، مشاهداتی که بیش از ٤۰٪ از شاخص‌های آن در دسترس نبود، از جامعه آماری حذف گردید. سایر مشاهدات گمشده، در صورت نرمال بودن با

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

جدول ۱. اطلاعات توصیفی از کلاس‌ها

	نام کلاس	جمعیت	میانگین بقای (ماه)
کلاس ۱	باقی‌مانده در روش دیالیز صفاقی	۱۰۳۹(٪۳۳/۵۵)	۳۹/٤۱±۲۷/۳۳
کلاس ۲	انتقال به همودیالیز	۵۸٤(٪۱۸/۸٦)	۲۵/٦٤±۲۲/۵٦
کلاس ۳	دریافت پیوند کلیه	۵٤۵(٪۱۷/٦۰)	۲۰/۵۵±۱۷/۱۲
کلاس ٤	فوت‌شده	۸۸۷(٪۲۸/٦٤)	۲٤/۳۷±۲۰/۵۸
کلاس ۵	بهبود عملکرد کلیه	٤۲(٪۱/۳٦)	۱۲/۲۳±۸/۰۹

میانگین حسابی و در غیر این صورت با مقدار میانه شاخص‌ها جایگذاری شده است. به‌منظور کاهش تاثیر بزرگی یا کوچکی مقادیر ویژگی‌ها و دست‌یابی به نتیجه دقیق‌تر، داده‌ها در بازه [۱ و ۰] مقیاس‌بندی شدند. درنهایت، ۷۵٪ از داده‌ها به‌منظور آموزش و مابقی برای آزمایش مدل مورد استفاده قرار گرفت.

پس از آماده‌سازی مجموعه داده‌ها، وضعیت بقای بیماران به پنج کلاس تقسیم شده‌اند. جدول ۱ داده‌های توصیفی کلاس‌ها را به‌صورت خلاصه نشان می‌دهد. میانگین مدت بقای بیماران در کلاس ۱ که مربوط به بیماران باقی‌مانده در روش دیالیز صفاقی سرپایی مداوم می‌باشد، بیشتر از سایر کلاس‌ها و در کلاس پنج که مربوط به بیماران دارای بازگشت عملکرد کار کلیه می‌باشد، کمترین مقدار است. همچنین حدود ۳۵٪ بیماران، روش دیالیز صفاقی سرپایی مداوم را به دلایلی مختلفی ترک کرده‌اند که مهمترین آن پریتونیت (٪۳۲/٤۵) و عوارض مکانیکی (٪۱۱/٤۳) بوده است.

الگوریتم رندوم فارست برای شناسایی مهمترین فاکتورها مورد استفاده قرار گرفت. با پیاده‌سازی این الگوریتم، عواملی که میزان اهمیت آن‌ها بیش از ۱۰ شد، به‌عنوان عوامل پیش‌بینی کننده بقای دیالیز صفاقی سرپایی مداوم مورد بررسی قرار گرفت. در جدول ۲، داده‌های مربوط به ۳۵ شاخص به کار گرفته‌شده در پیش‌بینی بقای بیماران، آورده شده است. فاکتورهای مدت زمان، وضعیت تغذیه، کفایت دیالیز، سن و حجم ادرار، به ترتیب جز مهمترین پارامترهای پیش‌بینی شناخته‌شده است.

جدول ۳ نتایج پیاده‌سازی الگوریتم‌ها را نشان می‌دهد. ابتدا الگوریتم ماشین بردار پشتیبان یکی در برابر همه برای طبقه‌بندی بقای بیماران انجام شد. با استفاده از این روش دقت پیش‌بینی بسیار کم به دست آمد. سپس الگوریتم ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری پس از پیاده‌سازی تابع تقسیم‌بندی، توسط استراتژی اعتبارسنجی k-fold انجام شد که در این‌جا از 10-fold به‌منظور افزایش دقت طبقه‌بندی استفاده شده و مجموعه داده کامل به‌صورت تصادفی به ۱۰ قسمت مساوی افراز و طبقه‌بندی ۱۰ مرتبه مورد آزمایش قرار گرفت. در نهایت دقت پیش‌بینی توسط روش ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری به‌طور قابل‌توجهی افزایش یافته و در حدود ۳۸٪ نسبت به روش ماشین بردار پشتیبان یکی در برابر همه بهبود دقت پیش‌بینی را نشان داد.

زمان آموزش ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری بیشتر از ماشین بردار پشتیبان یکی در برابر همه می‌باشد که این مسئله به‌دلیل استفاده از استراتژی اعتبار سنجی متقابل k-fold قابل توجیه است. این در حالی است که زمان آزمایش ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری کمتر از ماشین بردار پشتیبان یکی در برابر همه شد.

بحث

بیماری کلیوی مرحله نهایی، کاهش غیرقابل‌برگشت عملکرد کلیه است که در مواقع شدید، در صورت عدم انجام دیالیز یا پیوند کلیه، منجر به مرگ بیمار می‌گردد.[۱۳] این بیماری یک مشکل جهانی در حال رشد با پیامدهای اقتصادی است، که با وجود بهبود درمان‌های جایگزین کلیوی، خطر مرگ‌ومیر این بیماران بالا می‌باشد.[۱٤] برای درمان مرحله نهایی نارسایی کلیوی روش‌های متفاوتی وجود دارد.

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

جدول ۲: عوامل پیش‌بینی‌کننده بقای بیماران دیالیز صفاقی سرپایی مداوم

شاخص	میانگین±انحراف معیار	تعداد(درصد)	درجه اهمیت
تعداد بیماران	–	۳۰۹۷	–
جنسیت	–	–	–
مرد	–	۱۵۷۳(٪۵۰/۷۹)	۲۸/۵۶
زن	–	۱۵۲۴(٪۴۹/۲۱)	۲۷/۴۴
مدت زمان پیگیری (ماه)	۲۴/۴۸±۱۹/۱۳	–	۲۰۰/۵۵
سن (سال)	۵۰/۶۳±۱۵/۶۷	–	۱۱۰/۰۱
BMI (kg/m^2)	۲۳/۵۵±۴/۶۸	–	۴۷/۲۵
علت ESRD	–	–	–
دیابت ملیتوس	–	۱۲۶(٪۴۰/۹۱)	۲۵/۵۷
فشارخون	–	۸۶۶(٪۲۷/۹۷)	۲۲/۸۲
گلومرولونفریت	–	۱۸۶(٪۶/۰)	۱۴/۵۷
بیماری کلیه پلی کیستیک	–	۱۱۶(٪۳/۷۴)	۱۱/۸۱
سنگ کلیه	–	۷۹(٪۲/۵۵)	۱۲/۹۱
ناشناخته	–	۴۵۲(٪۱۴/۵۹)	۱۷/۹۹
سایر*	–	۱۳۱(٪۴/۲۳)	۱۰/۸۱
nPCR (g/kg/day)	۰/۸۶±۰/۳۴	–	۱۲۰/۶۲
هموگلوبین (mg/dl)	۱۰/۹۸±۲/۰۲	–	۲۲/۲۸
آلبومین (g/dl)	۳/۶۲±۰/۶۴	–	۱۶/۴۵
کراتینین(mg/dl)	۶/۹۴±۲/۹۷	–	۱۷/۸۴
کلسیم (mg/dl)	۸/۹۹±۱/۷۵	–	۲۳/۰۷
فسفر (mg/dl)	۴/۷۸±۱/۴۷	–	۲۶/۳۱
DBP (mmHg)	۸۰/۳۷±۱۳/۱۳	–	۵۲/۶۹
SBP (mmHg)	۱۳۳/۱۷±۲۳/۷۵	–	۶۸/۰۴
تری‌گلیسیرید (mg/dl)	۱۶۹/۳۴±۹۶/۳۰	–	۲۳/۵۹
LDL (mg/dl)	۱۱۳/۷۲±۴۳/۳۹	–	۲۲/۰۵
HDL (mg/dl)	۴۴/۴۹±۲۱/۳۳	–	۳۰/۶۸
کلسترول (mg/dl)	۱۹۰/۰۹±۵۲/۱۸	–	۳۱/۲۱
پتاسیم (mg/dl)	۴/۵۶±۰/۹۹	–	۲۳/۰۷
PTH (pg/dl)	۲۳۴/۳۱±۲۴۷/۴۲	–	۲۳/۷۳
فریتین (ng/dl)	۳۷۳/۱۲±۳۹۲/۴۱	–	۲۱/۹۴
Total KT/V	۱/۹۸±۰/۷۵	–	۱۱۷/۶۵
اسید اوریک (mg/dl)	۶/۲۳±۱/۹۴	–	۲۴/۱۹
نیتروژن اوره خون (mg/dl)	۶۶/۱۸±۳۷/۳۸	–	۱۷/۱۶
حجم ادرار (ml/day)	۷۶۹/۱۰±۵۸۴/۷۵	–	۱۰۹/۵۶
اولترافیلتراسیون (ml/day)	۹۳۳/۲۱±۵۳۸/۴۶	–	۲۷/۶۵
بیماری زمینه‌ای	–	–	–
بدون بیماری	–	۷۲۱(٪۲۳/۳۰)	۱۸/۲۳
فشارخون	–	۱۶۹۴(٪۵۴/۷۲)	۲۶/۲۷
دیابت ملیتوس	–	۹۲۷(٪۲۹/۹۵)	۲۳/۱۳
سایر**	–	۳۶۶(٪۱۱/۸۳)	۱۳/۸۵

* کلاژن و اسکولار، UTI مکرر، رفلاکس، بیماری‌های فامیلیال، توکسین- داروها/ ** بیماری کرونر، بیماری عروقی-مغزی، سرطان

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

جدول ۳: مقایسه دو روش از لحاظ دقت و زمان

	دقت پیش‌بینی	زمان آموزش (ثانیه)	زمان آزمایش (ثانیه)
OAA-SVM	۰/۵۱۹۹٤	٦/۱۱	۳/٤۸
MBT-SVM	۰/۸۹۵۷۸	۹/۵۲	۰/۷۳

دیالیز صفاقی یکی از شایع‌ترین روش‌های جایگزین کلیه برای بیماران مرحله نهایی کلیوی است که به‌منظور حفظ عملکرد کلیه‌های باقی‌مانده و کاهش بی‌ثباتی‌های قلبی-عروقی به‌کار می‌رود.[۱۵] این روش در مقایسه با همودیالیز، هزینه کمتری دارد و دارای نرخ بقای قابل مقایسه و کیفیت زندگی بهتری نسبت به همودیالیز می‌باشد.[۱۶و۱۷] اما متأسفانه گزارش‌های متعددی نشان داده است که درمان دیالیز صفاقی، پس از گذشت دو سال اول درمان، مطلوب‌تر از درمان همودیالیز نیست.[۱۸] از طرفی نسبت بیماران انتقال از دیالیز صفاقی به همودیالیز بیشتر از انتقال همودیالیز به دیالیز صفاقی می‌باشد.[۱۹] بنابراین بررسی بقای درازمدت بیماران دیالیز صفاقی، یکی از مهمترین چالش‌های سیستم‌های بهداشتی به‌شمار می‌رود.

برخی از مطالعات بالینی تلاش کردند تا عوامل مؤثر بر بقای بیماران دیالیز صفاقی سرپایی مداوم را بررسی کنند. از آن‌جایی‌که اهمیت نسبی این عوامل در بیماران مختلف متفاوت است، الگوریتم رندوم فارست به‌منظور شناسایی مهمترین شاخص‌ها به‌کار گرفته شد و اغلب عوامل تعیین‌شده، جز مهمترین عواملی هستند که در بسیاری از گزارش‌های دیگر مورد استفاده قرار گرفته است.[٦و۱۵و۲۰] همچنین با وجود پژوهش‌های انجام‌شده در زمینه بقای دیالیز صفاقی، بیشتر مطالعات به‌صورت توصیفی بوده و تعداد محدودی از آن‌ها از داده‌های بزرگ و چندمرکزی، با به‌کارگیری الگوریتم‌های ماشین یادگیری به تحلیل بقای بیماران دیالیز صفاقی سرپایی مداوم پرداخته‌اند. پیش‌بینی وضعیت آتی بیماران دیالیز صفاقی به‌صورت چند کلاسی، کمتر مورد ارزیابی قرار گرفته، که در این‌جا پنج کلاس برای بیماران دیالیز صفاقی سرپایی مداوم بررسی‌شده است. ماشین بردار پشتیبان، مدل‌های پیش‌بینی را با آموزش داده‌ها استنتاج می‌کند. مزیت اصلی آن دقت طبقه‌بندی آن می‌باشد. به‌طور کلی ماشین بردار پشتیبان در اغلب اوقات بهترین دقت طبقه‌بندی را ارایه می‌دهد.[۲۱] در کنار این مزیت یک محدودیت اساسی وجود دارد که برای مسائل طبقه‌بندی دوکلاسه طراحی شده است. برای رفع این مشکل، از روش‌های طبقه‌بندی چندکلاسه ماشین بردار پشتیبان با کاهش مسئله اصلی به مسائل باینری چندگانه، استفاده می‌گردد. به این منظور دو روش ماشین بردار پشتیبان یکی در برابر همه و ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری به‌کار گرفته شد. با پیاده‌سازی ماشین بردار پشتیبان یکی در برابر همه، دقت پیش‌بینی تعیین شده بسیار کم و نامناسب بوده که به دلیل عدم ایجاد مجموعه آموزشی متوازن و از بین رفتن تقارن مسئله می‌باشد. بنابراین، روش ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری که توانایی ایجاد مجموعه آموزشی متقارن‌تر، بر اساس تابع تقسیم‌بندی پراکندگی درون مجموعه‌ای را دارد، به‌کار گرفته شد. در این روش گاهی مشاهدات مجموعه داده‌های آزمایشی، ساده و یا سخت می‌شود، که در چنین مواردی دقت طبقه‌بندی قابل اعتماد نبوده و برای رفع این مشکل از اعتبارسنجی متقابل 10-fold استفاده شده است.[۲۲]

میانگین مدت‌زمان بقای بیماران ۱۹/۱۳±۲٤/٤۸ ماه بوده است. همچنین نسبت جنسیتی به این صورت است که ۱۵۷۳(۵۰/۷۹٪) مرد و ۱۵۲٤(٤۹/۲۱٪) زن می‌باشد که این نسبت در تمامی کلاس‌ها کمابیش مشابه بوده است. میانگین سنی افراد در سه کلاس ۱، ۲ و ۵ به‌صورت یکسان و با میانگین ۱٦/۷۹±۵۱/٦۲ سال می‌باشد. اما میانگین سنی در کلاس ۳ که مربوط به بیمارانی است که درنهایت پیوند کلیه می‌شوند، کاهش‌یافته و در حدود ۱۳/۲۹±۳۸/٤۸ سال و در کلاس ٤ بیمارانی که فوت شده‌اند، میانگین سنی افزایش یافته و ۱۷/٦۱±٦۱/۷۹ سال می‌باشد. Mujais و Story در مطالعه‌ای ثابت کردند هریک سال افزایش در سن، ٤٪ افزایش خطر مرگ‌ومیر را به‌دنبال دارد.[۲۳] سرم آلبومین و وضعیت تغذیه‌ای (Normalized

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

protein catabolic rate, nPCR) بیماران، در کلاس ٤ در کمترین وضعیت با میانگین‌های به ترتیب ۰/۵۸±۳/۳۳ و ۰/۲۳±۰/۷۷۷ می‌باشد. Hakemi و همکاران در گزارشی سرم آلبومین و وضعیت تغذیه را از عوامل پیش‌بینی کننده مرگ‌ومیر دانستند و نشان دادند حدود ۸۰٪ از بیماران وضعیت تغذیه‌ای کمتر از استاندارد و ٤۰٪ بیماران آلبومین کمتر از mg/dl ۳/۵ دارند.[۲٤] افزون‌بر فاکتورهای بررسی‌شده، هورمون پاراتیرویید (Parathyroid hormone, PTH) یک عامل پیشگیری‌کننده مستقل بقای بیمار دیالیز صفاقی سرپایی مداوم ارزیابی‌شده است، به‌طوری‌که خطر مرگ بیماران با سطح هورمون پاراتیرویید پایین‌تر، افزایش می‌یابد.[۲۵] میانگین هورمون پاراتیرویید در کلاس ۱ بیشترین مقدار برابر ۲٦٦/۱۷±۲۸۹/٦۲ و در کلاس ٤ کمترین مقدار برابر ۲۳۲/۵۷±۱۷۷/۸۹ را دارد. میانگین سطح سرمی فریتین در کلاس ٤ برابر ٤٦۱/۱۹±٤۹۷/۵۵ نشان داده‌شده که تفاوت قابل‌توجهی با میانگین سایر کلاس‌ها دارد، که می‌تواند روی بقای بیماران مؤثر باشد. Yousefi و همکارشان نشان دادند که سطح بالای فریتین به‌عنوان یکی از عوامل افزایش خطر مرگ‌ومیر بیماران دیالیز صفاقی است.[۲٦]

متوسط حجم ادرار روزانه، در کلاس‌های مختلف به‌صورت مشابه و در حدود ٦۱۵/۷٤±۸۳۷/٦۰ بوده است، اما در کلاس ٤ کمترین مقدار را با میانگین ۵۲۱/٦٦±٦۰۷/۲۹ به خود اختصاص داده بود. بر اساس گزارشات، بیمارانی که حجم ادرار ۲٤ ساعته آن‌ها کمتر از ml ۲۵۰ بوده، دو برابر بیشتر از کسانی که حجم ادرار بیشتری داشته‌اند، در معرض خطر مرگ‌ومیر بوده‌اند.[۲٤] در سایر پارامترها برآورد میانگین‌های عوامل در کلاس‌های مختلف کمابیش یکسان و تفاوت معناداری بین کلاس‌ها وجود نداشت.

شایعترین بیماری زمینه‌ای، فشارخون با ٥٤/۷۲٪ بوده است که در تمامی کلاس‌ها تا حدی یکسان است. پس از آن بیماری دیابت ملیتوس به‌عنوان بیشترین درصد بیماری همراه با ۲۹/۹۵٪ تشخیص داده شد. بیماری دیابت ملیتوس در کلاس ۳ کمترین مقدار (٦/۸٪) و در کلاس ٤ بیشترین مقدار (۳۱/۸٪) را به خود اختصاص داده بود. همچنین ۲۳/۳٪ از بیماران، فاقد بیماری زمینه‌ای بوده‌اند که این عدد در کلاس ۵، یعنی بیمارانی که بازگشت عملکرد کلیه داشته‌اند، حدود ۳٦٪ بوده است. باوجود این‌که شایعترین بیماری همراه فشارخون بوده است، اما اصلی‌ترین علت نارسایی کلیه بیماری دیابت شناخته شد. ٤۰/۸۸٪ از بیماران به‌علت بیماری دیابت ملیتوس دچار نارسایی کلیه شده‌اند. این درصد در بیمارانی که در کلاس ۳ و ۵ هستند بسیار کمتر و در حدود ۱٦/۵٪ می‌باشد، درحالی‌که در کلاس ٤ این عدد به ۵۷/۱٤٪ رسید. یعنی ۵۷/۱٤٪ از بیمارانی که در نهایت فوت شده‌اند علت اصلی نارسایی کلیه‌شان، دیابت بوده است. Mujais و همکارشان نشان دادند بیماران مبتلا به دیابت در مقایسه با بیماران غیردیابتی تا ۳۰٪ افزایش خطر مرگ‌ومیر را دارند.[۲۳] افزون‌براین دلایل متعددی باعث مرگ بیماران دیالیز صفاقی سرپایی مداوم می‌شود که در بین آن بیماری قلبی با ۵۲/۸۸٪ اصلی‌ترین عامل، ۱۱/٦٪ بیماران به دلیل سکته مغزی و ۷/۵٪ بیماران به دلایل عفونی فوت شده‌اند. گزارش‌هایی مشابه نیز وجود داشت. Buraczynska و همکاران دریافتند بیماری‌های قلبی–عروقی با حدود ۵۰٪، دلیل اصلی مرگ بیماران دیالیز صفاقی می‌باشد.[۲۷] همچنین Shlipak و همکاران، سکته مغزی با حدود ۱۰٪ علت دوم یا سوم مرگ‌ومیر را در طول درمان دیالیز صفاقی سرپایی مداوم دانستند.[۲۸]

در نهایت، تجزیه‌وتحلیل نشان داد که سن بالا، سطح سرم آلبومین پایین، وضعیت تغذیه‌ای نامناسب، سطح هورمون پاراتیرویید پایین، سطح بالا سرم فریتین، حجم ادرار کم و علت نارسایی کلیوی دیابت، تأثیر منفی روی بقای بیماران دیالیز صفاقی دارد و احتمال مرگ این بیماران را افزایش می‌دهد. افرادی که علت نارسایی کلیه آن‌ها دیابت ملیتوس نبوده، سطح هورمون پاراتیرویید بالا و حجم ادرار بیشتر از ml ۷۰۰ در ۲٤ ساعت را داشته‌اند، کسانی هستند که امکان بازگشت عملکرد کلیه برای آن‌ها به‌عنوان عوامل اصلی سنجیده شده است. همچنین بیمارانی با فاکتورهای اصلی سن پایین و نداشتن بیماری همراه، افرادی هستند که با احتمال بیشتری تحت پیوند کلیه قرار می‌گیرند. افزون‌براین با توجه به پراکندگی و حجم مجموعه داده‌ها، می‌توان دریافت که ماشین بردار پشتیبان چندکلاسه، روشی مناسب برای پیش‌بینی بقای بیماران دیالیز صفاقی سرپایی مداوم است. همچنین با مقایسه دو روش ماشین بردار پشتیبان یکی در برابر همه و ماشین بردار پشتیبان مبتنی بر رویکرد نگاشت چند فضایی درخت باینری می‌توان نتیجه گرفت که با به‌کارگیری تابع پارتیشن‌بندی مناسب و استفاده از اعتبارسنجی متقابل k-fold، درستی پیش‌بینی ماشین‌های بردار پشتیبان چند کلاسه را می‌توان افزایش داد. بنابراین روش ماشین‌بردار پشتیبان مبتنی بر رویکرد نگاشت چند

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

فضایی درخت باینری با توجه به شاخص‌های ارزیابی متعدد و پراکندگی در توزیع کلاس‌ها، می‌تواند با دقت پیش‌بینی بالایی نسبت به روش ماشین بردار پشتیبان یکی در برابر همه وضعیت بیماران دیالیز صفاقی سرپایی مداوم را مورد ارزیابی قرار دهد. در این پژوهش، تابع تقسیم‌بندی براساس پراکندگی درون مجموعه به‌کار گرفته شده است. پیشنهاد می‌شود از توابع دیگر پارتیشن‌بندی از جمله تابع شباهت گروهی، تابع تقسیم‌بندی مبتنی بر مرکز مسئله و یا ترکیبی از این توابع استفاده شود. همچنین این روش قابلیت پیش‌بینی بقای روش درمان همودیالیز و حتی سایر بیماری‌ها را دارد. از این‌رو می‌توان آن را برای هر نوع بیماری دیگر مورد استفاده قرار داد.

References

1. Heaf J. High transport and malnutrition-inflammation-atherosclerosis (MIA) syndrome. *Perit Dial Int* 2003 23(2):109-10.
2. Oreopoulos DG, Robson M, Faller B, Ogilvie R, Rapoport A, deVeber GA. Continuous ambulatory peritoneal dialysis: a new era in the treatment of chronic renal failure. *Clin Nephrol* 1979;11(3):125-8.
3. Kim DK, Lee SM, Son YK, Kim SE, Kim KH, An WS. Factors influencing survival according to elapsed time in peritoneal dialysis patients. *Ren Fail* 2012;34(5):559-65.
4. Peng YS, Chiang CK, Hung KY, Chang CH, Lin CY, Yang CS, et al. Comparison of self-reported health-related quality of life between Taiwan hemodialysis and peritoneal dialysis patients: a multi-center collaborative study. *Qual Life Res* 2011;20(3):399-405.
5. Zhang F, Liu H, Gong X, Liu F, Peng Y, Cheng M, et al. Risk factors for mortality in Chinese patients on continuous ambulatory peritoneal dialysis. *Perit Dial Int* 2015;35(2):199-205.
6. Kendrick J, Teitelbaum I. Strategies for improving long-term survival in peritoneal dialysis patients. *Clin J Am Soc Nephrol* 2010;5(6):1123-31.
7. Tesfaye B, Atique S, Elias N, Dibaba L, Shabbir SA, Kebede M. Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach. *Comput Methods Programs Biomed* 2017;140:45-51.
8. Noori S, Nourijelyani K, Mohammad K, Niknam MH, Mahmoudi M, Andonian L, Akaberi A. Random forests analysis: A modern statistical method for screening in high-dimensional studies and its application in a population-based genetic association study. *J North Khorasan Univ Med Sci* 2012;3(5):93-101.
9. Wang Z, Xue X. Multi-class support vector machine. Ma Y, Guo G, editors. Support Vector Machines Applications. New York, NY: Springer International Publishing; 2014. P. 23-48.
10. Safaee P, Noorossana R, Heidari K, Soleimani P. Using decision tree to predict serum ferritin level in women with anemia. *Tehran Univ Med J* 2016;74(1):50-7.
11. He X, Wang Z, Jin C, Zheng Y, Xue X. A simplified multi-class support vector machine with reduced dual optimization. *Pattern Recognit Lett* 2012;33(1):71-82.
12. Yang X, Yu Q, He L, Guo T. The one-against-all partition based binary tree support vector machine algorithms for multi-class classification. *Neurocomputing* 2013;113:1-7.
13. Jahantigh FF, Malmir B, Avilaq BA. A computer-aided diagnostic system for kidney disease. *Kidney Res Clin Pract* 2017;36(1):29-38.
14. Lin TH, Chuang SY, Chu CY, Lee WH, Hsu PC, Su HM, et al. The impact of chronic kidney disease on lipid management and goal attainment in patients with atherosclerosis diseases in *Taiwan. Int J Med Sci* 2014;11(4):381-8.
15. Jotterand Drepper V, Kihm LP, Kälble F, Diekmann C, Seckinger J, Sommerer C, et al. Overhydration is a strong predictor of mortality in peritoneal dialysis patients - independently of cardiac failure. *PLoS One* 2016;11(7):e0158741.
16. Klarenbach SW, Tonelli M, Chui B, Manns BJ. Economic evaluation of dialysis therapies. *Nat Rev Nephrol* 2014;10(11):644-52.
17. Weinhandl ED, Foley RN, Gilbertson DT, Arneson TJ, Snyder JJ, Collins AJ. Propensity-matched mortality comparison of incident hemodialysis and peritoneal dialysis patients. *J Am Soc Nephrol* 2010;21(3):499-506.
18. Raina AF, Saleem SM, Jan SS, Yousuf S. Mechanical complications of continuous ambulatory peritoneal dialysis: A hospital based retrospective study in Kashmir Valley. *Arch Med Health Sci* 2017;5(1):21-3.
19. Gulcan E, Kıdır V, Keles M, Cankaya E, Uyanik A, Saatci F. Factors affecting patient survival and technical survival in patients undergoing peritoneal dialysis. *Int J Clin Exp Med* 2017;10(1):1004-14.
20. Zhou Q, Zhou H, Li T. Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features. *Knowl-Based Syst* 2016;95:1-11.
21. Meyer D, Leisch F, Hornik K. The support vector machine under test. *Neurocomputing* 2003;55(1):169-86.
22. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, et al. Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst* 2012;36(4):2431-48.
23. Mujais S, Story K. Peritoneal dialysis in the US: evaluation of outcomes in contemporary cohorts. *Kidney Int Suppl* 2006;(103):S21-6.
24. Hakemi MS, Golbabaei M, Nassiri A, Kayedi M, Hosseini M, Atabak S, et al. Predictors of patient survival in continuous ambulatory peritoneal dialysis: 10-year experience in 2 major centers in Tehran. *Iran J Kidney Dis* 2010;4(1):44-9.
25. Hashemi J, Hesari Z, Golshan A. Evaluation of calcium, phosphorus and potassium in saliva and their relationship to blood biochemical factors in hemodialysis patients. *Tehran Univ Med J* 2017;75(1):65-71.
26. Yousefi Abdolmalehi E, Seyfi S. Although maintenance dialysis prevents death from uremia, patient survival and quality of life remains an important issue. *J Mazandaran Univ Med Sci* 2012;21(86):213-20.
27. Buraczynska M, Ksiazek P, Zukowski P, Benedyk-Lorens E, Orlowska-Kowalik G. Complement factor H gene polymorphism and risk of cardiovascular disease in end-stage renal disease patients. *Clin Immunol* 2009;132(2):285-90.
28. 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.

[Downloaded from tumj.tums.ac.ir on 2025-04-04]

TUMJ Tehran University Medical Journal, January 2018; Vol. 75, No. 10: 752-760 Original Article

A model for predicting peritoneal dialysis patients' survival, using data mining algorithms

Farzad Firouzi Jahantigh Ph.D.[1*]
Iraj Najafi Ph.D.[2]
Maryam Ostovare M.Sc.[1]

1- Department of Industrial Engineering, Faculty of Engineering Shahid Nikbakht, Sistan and Baluchestan University, Zahedan, Iran.
2- Department of Nephrology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran.

[*]Corresponding author: Department of Industrial Engineering, Faculty of Engineering, Sistan and Baluchestan University, University Blvd., Sistan and Baluchestan, Zahedan, Iran.
Tel: +98- 21- 88905011- 2
E-mail: Firouzi@eng.usb.ac.ir

Abstract

Received: 30 Aug. 2017 Revised: 30 Dec. 2017 Accepted: 04 Jan. 2018 Available online: 05 Jan. 2018

Background: Peritoneal dialysis is one of the most commonly used treatment methods for the patients with end stage renal failure. In recent years, the mortality rate of patients under this treatment has decreased; however, long-term survival is still an important challenge for health systems. The present study aimed to predict the survival of continuous ambulatory peritoneal dialysis patients.

Methods: In this retrospective study, according to the difference of relative importance of demographic characteristics, laboratory data, dialysis adequacy parameters and nutritional status in various patients, the factors affecting the survival of peritoneal dialysis patients have been identified by random forest algorithm. Then, the clinical and laboratory data of patients undergoing continuous ambulatory peritoneal dialysis treatment were evaluated retrospectively from July 1996 to April 2014 in 18 peritoneal dialysis centers, using multi-class one against all support vector machine (OAA-SVM) and multi-space mapped binary tree support vector machine (MBT-SVM) algorithms.

Results: 3097 patients were studied with the mean age of 50.63±15.67 years and average follow-up time of 24.48±19.13 months. The results of the random forest algorithm have identified 35 factors as the most important predictors of peritoneal dialysis patient's survival. Then, the prediction of peritoneal dialysis patients' survival status was evaluated using one against all support vector machine and multi-space mapped binary tree support vector machine algorithms in 5 classes of patients including "still on peritoneal dialysis", "transferred to hemodialysis", "received a kidney transplant", "died" and "improved kidney function". The reliability of survival prediction algorithms were 51.99% and 89.57% respectively.

Conclusion: An accurate prediction model would be a potentially useful way to evaluate patients' survival at peritoneal dialysis that increased clinical scrutiny and timely intervention could be brought to bear. So, in this research, the multi-space mapped binary tree support vector machine algorithm has a high precision in predicting the survival of continuous ambulatory peritoneal dialysis patients considering multiple evaluation indices and different class distribution functions.

Keywords: continuous ambulatory peritoneal dialysis, data mining, health status, retrospective studies, survival rate.

[Downloaded from tumj.tums.ac.ir on 2025-04-04]