

Original Article

Prediction of Urinary Tract Infection for Hospital-admitted Patients based on Demographic and Historical Data, as well as Machine Learning Approaches

Sajjad Farashi^{1,2*} , Hossein Emad Momtaz³ 

1. Neuroscience and Mental Health Research Institute, Avicenna Health Research Institute, Hamadan University of Medical Sciences, Hamadan, Iran
2. Urology and Nephrology Research Center, Avicenna Clinical Sciences Research Institute, Avicenna Health Research Institute, Hamadan University of Medical Sciences, Hamadan, Iran
3. Department of Pediatrics, Faculty of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran

Article history:

Received: 16 June 2024

Revised: 03 December 2024

Accepted: 11 December 2024

***Corresponding author:** Sajjad Farashi, Neuroscience and Mental Health Research Institute, Avicenna Health Research Institute, Hamadan University of Medical Sciences, Hamadan, Iran

Email: sajjad_farashi@yahoo.com

Abstract

Background and Objective: Urinary tract infection (UTI) is one of the common infections that affects the urinary system. UTI can be detected using analysis of urine culture which is a time-consuming and error-prone procedure. The fast prediction of UTI helps to start antibiotic medication at the correct time and prior to the culture report. The present study aimed to assess the potential of artificial intelligence for UTI prediction.

Materials and Methods: The current study was conducted based on the published data from a retrospective cohort study that was performed on 300,000 human samples in Denmark. The performance of machine learning algorithms, including support vector machines, decision trees, linear discriminant analysis, and linear regression models, was investigated for UTI prediction according to historical and demographical features.

Results: The obtained results pointed out that UTI prediction using a linear discriminant analysis model with an accuracy of $65.16 \pm 0.64\%$ was possible. Among the features, age had a significant effect on UTI prediction where the model accuracy rates for elderly, adult, young, and infant cases were reported as 73.64, 86.25, 39.42, and, 60.60%, respectively. In addition, the results demonstrated that the classification performance was better for adults compared to younger participants. The most informative features were also suggested according to the neighborhood component analysis.

Conclusion: The results of this study highlighted the potential of artificial intelligence for fast UTI prediction; nonetheless, the performance needs to be enhanced by the addition of other fast-accessible features.

Keywords: Infection, Machine learning, Prognosis, Urinary tract

Please cite this article as follows: Farashi S. Prediction of Urinary Tract Infection for Hospital-admitted Patients based on Demographic and Historical Data, as well as Machine Learning Approaches. J Res Urol 2024; 8(1):18-26. DOI: 10.32592/jru.8.1.18



پیش بینی عفونت ادراری در بیماران بستری در بیمارستان براساس داده‌های دموگرافیک و سوابق پزشکی مبتنی بر رویکردهای یادگیری ماشین

سجاد فراشی^{۱،۲*}، حسین عماد ممتاز^۳

۱. پژوهشکده علوم اعصاب و سلامت روان، پژوهشگاه سلامت ابن سینا، دانشگاه علوم پزشکی همدان، همدان، ایران
۲. مرکز تحقیقات اورولوژی و نفرولوژی، پژوهشکده علوم بالینی ابن سینا، پژوهشگاه سلامت ابن سینا، دانشگاه علوم پزشکی همدان، همدان، ایران
۳. گروه اطفال، دانشکده پزشکی، دانشگاه علوم پزشکی همدان، همدان، ایران

چکیده

سابقه و هدف: عفونت ادراری (UTI) یکی از عفونت‌های شایع است که سیستم ادراری را تحت تأثیر قرار می‌دهد. روش تشخیص استاندارد UTI تجزیه و تحلیل کشت ادرار است. با این حال، این روش تشخیصی زمان‌بر بوده و به شدت مستعد خطاست. با توجه به اینکه پیش‌بینی سریع عفونت ادراری به شروع به موقع مصرف آنتی‌بیوتیک و کنترل بهینه تر عفونت ادراری کمک می‌کند، توسعه روش‌های تشخیص سریع تر مورد توجه محققان این حوزه قرار دارد.

مواد و روش‌ها: مطالعه حاضر براساس داده‌های منتشر شده یک مطالعه کوهورت گذشته‌نگر که در کشور دانمارک و با حجم نمونه ۳۰۰۰۰۰ به انجام رسیده، انجام شده است. در مطالعه حاضر و براساس داده موجود، توانمندی الگوریتم‌های یادگیری ماشین شامل بردارهای پشتیبان، درخت تصمیم‌گیر، آنالیز تفکیک‌کننده خطی و ماشین رگرسیون خطی در پیش‌بینی عفونت ادراری براساس پیش‌گوکننده‌هایی نظیر سوابق پزشکی و داده‌های دموگرافیک برای بیماران بستری در بیمارستان ارزیابی شده است.

یافته‌ها: نتایج این مطالعه نشان داد که طبقه‌بندی تفکیک‌کننده خطی با دقت $65/16 \pm 0/64$ پیش‌بینی UTI را با استفاده از ویژگی‌های دموگرافیک و سوابق پزشکی بیمار انجام می‌دهد. از میان ویژگی‌ها، سن تأثیر قابل توجهی بر پیش‌بینی UTI داشت؛ به نحوی که دقت مدل در تشخیص UTI در نمونه‌های کهن‌سال، بزرگ‌سال، کودک و جوان به ترتیب $73/64$ ، $86/25$ و $39/42$ بوده است. ویژگی‌های با قدرت تمایز بالا نیز با توجه به تجزیه و تحلیل مؤلفه‌های همسایگی پیشنهاد شد.

نتیجه‌گیری: نتایج این مطالعه پتانسیل هوش مصنوعی را برای پیش‌بینی سریع UTI نشان می‌دهد. با این حال، عملکرد سیستم پیش‌بینی‌کننده با افزودن ویژگی‌های دیگر قابل افزایش خواهد بود.

واژگان کلیدی: دستگاه ادراری، عفونت، پیش‌بینی، یادگیری ماشین

تاریخ دریافت مقاله: ۱۴۰۳/۰۳/۲۷

تاریخ ویرایش مقاله: ۱۴۰۳/۰۹/۱۳

تاریخ پذیرش مقاله: ۱۴۰۳/۰۹/۲۱

تمامی حقوق نشر برای دانشگاه علوم پزشکی همدان محفوظ است.

* نویسنده مسئول: سجاد فراشی،

پژوهشکده علوم اعصاب و سلامت روان، پژوهشگاه سلامت ابن سینا، دانشگاه علوم پزشکی همدان، همدان، ایران

ایمیل: sajjad_farashi@yahoo.com

استناد: فراشی، سجاد. پیش‌بینی عفونت ادراری در بیماران بستری در بیمارستان براساس داده‌های دموگرافیک و سوابق پزشکی مبتنی بر رویکردهای یادگیری ماشین. تحقیقات در ارولوژی، بهار و تابستان ۱۴۰۳: ۱۸ (۱): ۲۶-۱۸

مقدمه

زندگی خود تجربه می‌کنند [۱]. عفونت‌های ادراری همچنین یکی از علل اصلی مرگ‌ومیر در دنیا هستند. UTI از نظر بالینی براساس علائم بالینی و تجزیه و تحلیل کشت ادرار تشخیص داده می‌شود. با این حال، راهبردهای تشخیص‌دهنده فعلی کمتر از حد مد نظر

عفونت‌های دستگاه ادراری (Urinary tract infection (UTI) عفونت‌های شایعی هستند که چندین قسمت از دستگاه ادراری از جمله کلیه‌ها و مثانه را تحت تأثیر قرار می‌دهند. براساس گزارش‌ها، ۵۰ درصد از زنان حداقل یک‌بار عفونت ادراری را در طول

بهینه و همچنین زمان بر هستند [۲]. زمان طولانی مورد نیاز برای در دسترس بودن جواب آزمایش‌های کشت ادرار (حدود ۴۸ تا ۷۲ ساعت) و همچنین روش‌های مورد استفاده برای تهیه نمونه ادرار که بسیار مستعد آلوده شدن نمونه به آلودگی‌های محیطی هستند، از جمله موارد مشکل‌ساز در تشخیص‌های فعلی عفونت‌های ادراری‌اند. علاوه بر این، روش‌های تشخیص بالینی فعلی نیز تا ۳۳ درصد مستعد خطا هستند [۳]. این مسائل ضرورت توسعه روش‌های تشخیصی مکمل را برجسته می‌کنند.

چندین عامل از جمله عفونت‌های ادراری قبلی، سن [۴]، دیابت [۵]، قرار گرفتن در معرض محرک‌های محیطی و عوامل ژنتیکی [۶]، فعالیت جنسی و بی‌اختیاری ادرار [۷] از ریسک فاکتورهای بالقوه عفونت ادراری هستند. علاوه بر این، علائم عصبی حاد ممکن است با UTI در ارتباط باشند [۸]. از این رو، براساس شواهد موجود می‌توان انتظار داشت که مشخصه‌های دموگرافیک و سوابق پزشکی حاوی اطلاعات ارزشمندی برای پیش‌بینی احتمال UTI در زمان فعلی یا آینده باشند.

هوش مصنوعی (AI) یکی از استراتژی‌های ممکن برای تشخیص و پیش‌بینی UTI است [۹]. هوش مصنوعی با استفاده از الگوریتم‌های یادگیری ماشین سعی می‌کند از برنامه‌های کامپیوتری برای تجزیه و تحلیل داده‌های به دست آمده از ادرار انسان (یعنی داده‌های آزمایشگاهی) و داده‌های مشاهده‌ای، داده‌های سوابق پزشکی و غیره استفاده کند تا الگوهای پنهان مرتبط با سلامت دستگاه ادراری را از این دست داده‌ها شناسایی کند. هدف اصلی هوش مصنوعی، یافتن مهم‌ترین عوامل مخدوش‌کننده، یافتن رابطه و تعامل بین عوامل و در نتیجه شناسایی موارد فعلی ریسک یا پیش‌بینی نتایج آینده است. به طور خلاصه، هوش مصنوعی در غربالگری سلامت ادرار برای استخراج قوانین مفید برای تشخیص، پیش‌بینی و ارزیابی درمان استفاده می‌شود [۱۰].

برتون و همکاران (۲۰۱۹) از هوش مصنوعی برای یافتن مهم‌ترین مشخصه‌های نمونه‌های ادرار جهت پیش‌بینی ضرورت کشت باکتریایی استفاده کردند [۱۱]. این مطالعه نشان داد که رویکرد هوش مصنوعی ممکن است به کاهش ۳۹/۱ درصدی در لزوم کشت ادرار در تشخیص UTI منجر شود. گروهی از محققان مدلی را برای تشخیص کودکان با خطر عفونت ادراری بالا ایجاد کردند. این مدل متغیرهایی مانند سن، جنسیت، نژاد، فشارخون، سوزش ادرار، مواجهه قبلی با آنتی‌بیوتیک‌ها و نسبت آلبومین و کراتینین ادرار را دربر می‌گرفت. سطح زیرمنحنی مشخصه عملکرد برای کار طبقه‌بندی (شناسایی کودکان در معرض خطر) برای مطالعه آن‌ها ۰/۷۶ بود. این مدل می‌توانست مداخله پزشکی را برای کودکان UTI بهینه‌تر کند [۱۲]. اوزکان و همکاران (۲۰۱۸) با استفاده از معاینه معمول ادرار و تصاویر سونوگرافی پزشکی از سیستم ادراری و استفاده از چندین ابزار یادگیری ماشین (شامل شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان، درخت‌های تصمیم‌گیر و جنگل تصادفی) سیستمی را برای تشخیص UTI

پیشنهاد کردند [۱۰]. چنین سیستمی با استفاده از شبکه عصبی مصنوعی دقت طبقه‌بندی ۹۸/۳٪ را به دست می‌داد. گادالا Gadalla و همکاران (۲۰۱۹) به منظور تشخیص UTI از الگوریتم‌های یادگیری ماشین برای یافتن نشانگرهای زیستی ایمونولوژیک در نمونه‌های ادرار استفاده کردند. براساس مطالعه نان، با توجه به نتایج یادگیری ماشین، این امکان وجود دارد که بتوان برای موارد آسیب‌پذیر، راهکار درمانی مشخصی را ارائه کرد و از این طریق مخاطرات درمان را کاهش داد [۱۳]. جنگ Jeng و همکاران (۲۰۲۲) مدل‌های یادگیری ماشین را برای پیش‌بینی UTI عودکننده پیشنهاد کردند [۱۴]. آنان یک مدل جنگل تصادفی را با دقت ۷۰ درصد برای پیش‌بینی پیشنهاد دادند. دونگ و همکاران (۲۰۲۲) عملکرد تشخیصی ترکیب AI و میکرو تصویربرداری جریان را برای غربالگری سریع UTI بررسی کردند [۱۵]. نتایج این مطالعه توانمندی هوش مصنوعی را برای شناسایی سریع UTI نشان داد. اگرچه تاکنون چندین مطالعه برای توسعه ماشین‌های هوشمند جهت پیش‌بینی UTI انجام شده است، با توجه به پایین بودن دقت روش‌های پیشنهادی و یا دشواری در دسترس بودن داده‌های مورد نیاز، همچنان نیاز به انجام مطالعات بیشتر وجود دارد.

در مطالعه حاضر، رهیافت یادگیری ماشین برای مجموعه‌ای از داده‌های آنلاین UTI که در دسترس عموم قرار دارند، مورد استفاده قرار گرفته است. به صورت ویژه، ارتباط بین برخی از فاکتورهای موارد مشکوک به عفونت ادراری نظیر سن، جنسیت، سابقه عفونت و سابقه سایر بیماری‌ها با خطر عفونت ادراری بررسی شده است. علاوه بر این، کاربرد این دسته از فاکتورها برای تعیین حساسیت یک بیمار بستری در بیمارستان به UTI قبل از در دسترس بودن نتیجه کشت ادرار بررسی شده است. فرض این مطالعه بر این است که عوامل ذکر شده در بالا امکان توسعه یک مدل (ماشین هوشمند) برای پیش‌بینی UTI پیش از در دسترس قرار گرفتن جواب آزمایش کشت ادرار را به دست می‌دهند.

روش کار

توصیف داده

مجموعه داده‌های مورد استفاده در مطالعه حاضر داده‌های در دسترس عمومی هستند که از طریق لینک doi.org/10.5061/dryad.6djh9w107 قابل دسترس‌اند [۱۶]. این داده مربوط به بیماران بستری شده با علائم UTI در چهار بیمارستان در جنوب دانمارک است. برای بیماران بستری شده، از نتایج کشت ادرار برای برچسب زدن هر بیمار به عنوان مبتلا به UTI یا سالم (گروه شاهد یا غیرمبتلا) استفاده شده است. برای هر بیمار، داده‌هایی از جمله تاریخ و محل بستری، اطلاعات دموگرافیک (جنس، سن و شناسه بیمار)، کدهای تشخیص سوابق بیماری (بیماری انسدادی مزمن ریه، احتباس ادرار و بیماری عصبی)، تعداد UTI های قبلی، کاتتر ادراری ساکن (IUC) و نوع بستری (حاد یا

هسته خطی استفاده شد؛ زیرا مشاهدات ما نشان داد که این هسته بهترین نتیجه را به دست می‌دهد. برای طبقه‌بندی KNN، تعداد همسایگی ۱۰۰ و برای طبقه‌بند تفکیک‌کننده خطی از کرنل خطی استفاده شد. برای یادگیری، اعتبارسنجی و آزمایش مدل، مجموعه داده با استفاده از روش نمونه‌گیری تصادفی بدون جایگزینی به داده‌های آموزش، تست و اعتبارسنجی تقسیم شد. عملکرد طبقه‌بند آموزش‌دیده با استفاده از داده‌های آموزش براساس داده‌های اعتبارسنجی و مجموعه داده‌های آزمایشی ارزیابی شد. هر طبقه‌بند توسط یک مجموعه داده (۹۰٪ از کل مجموعه داده آموزش) آموزش داده و با استفاده از مجموعه داده دیگر (۱۰٪ از کل مجموعه داده آموزش) اعتبارسنجی گردید. به منظور جلوگیری از بیش‌برازش مدل، از روش اعتبارسنجی متقابل K-fold (K=10) استفاده شده است. عملکرد طبقه‌بند برای پیش‌بینی UTI با استفاده از داده‌های تست با حساسیت، ویژگی و دقت به ترتیب با استفاده از روابط ۱-۳ ارزیابی شد:

$$SE = \frac{TP}{TP+FN} \quad (۱)$$

$$SP = \frac{TN}{TN+FP} \quad (۲)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (۳)$$

که در آن TP، FN، TN و FP به ترتیب نشان‌دهنده مثبت درست، منفی کاذب، منفی درست و مثبت کاذب بودند. قابل ذکر است که بخش بزرگی از مجموعه داده (۹۳/۷۴ درصد) مربوط به موارد غیرمبتلا بود. این مسئله سبب نامتعادل شدن و بالانس نبودن مجموعه داده شده است. با توجه به بزرگ‌تر بودن حجم نمونه در گروه سالم، به منظور نزدیک کردن حجم داده‌ها در دو گروه سالم و بیمار، از روش down-sampling با انتخاب تصادفی نمونه‌ها استفاده شده است.

آنالیز آماری

برای مقایسه ویژگی‌های گروه‌های مبتلا و سالم (شاهد)، از آزمون کای ۲ استفاده شد. فرضیه صفر این بود که دو گروه با در نظر گرفتن یک ویژگی خاص متفاوت هستند. علاوه بر این، ارتباط بین متغیرها با استفاده از آزمون کوکران-مانتل-هنزل [۱۸] ارزیابی شد. اگر مقدار p کوچک‌تر از ۰/۰۵ باشد، فرضیه ارتباط بین متغیرها رد می‌شود. رابطه خطی بین متغیرها با استفاده از هم‌بستگی پیرسون و chi-squared p-value محاسبه شد.

یافته‌ها

آنالیز آماری داده

در مجموعه داده اصلی، ۶/۲۶٪ موارد UTI مثبت بودند، درحالی‌که سایر داده‌ها (۹۳/۷۴٪) نمونه‌های UTI منفی بودند.

برنامه‌ریزی‌شده) ثبت شده است. کدهای تشخیصی، متغیرهای طبقه‌بندی هستند (بله (۱) یا خیر (۰))، درحالی‌که تعداد UTI‌های قبلی، کاترهای ادراری ساکن (IUC) و همچنین متغیر سن، متغیرهای عددی هستند. این متغیرها به عنوان عوامل خطر برای پیش‌بینی UTI در آینده در نظر گرفته شده‌اند. حجم نمونه این مجموعه داده ۳۰۱۹۰۳ (۵۳/۸۵ درصد زن) و میانگین سن شرکت‌کنندگان ۵۰/۰۴±۲۶/۹۰ سال است. داده‌ها در بازه ژانویه ۲۰۱۷ تا آوریل ۲۰۱۸ و می ۲۰۱۸ جمع‌آوری شده‌اند. بازه زمانی ثبت داده‌ها (بیش از یک سال) تضمین کرده است که نتایج تجزیه و تحلیل‌ها نسبت به تغییرات فصلی بیماری‌های عفونت دستگاه ادراری حساس نباشد [۱۶].

فضای ویژگی

مجموعه داده استفاده شده در این مطالعه شامل اطلاعات متفاوتی برای هر شرکت‌کننده است که از آن جمله می‌توان به سن شرکت‌کننده، جنسیت (مرد یا زن)، بستری مجدد (درست یا نادرست)، وجود بیماری مزمن انسدادی ریه (COPD)، احتباس ادرار، بیماری عصبی، تعداد عفونت‌های دستگاه ادراری مرتبط با بیمارستان قبلی (previous_HAI_UTI)، عفونت‌های دستگاه ادراری مرتبط با جامعه (previous_CAI_UTI) و تعداد کاتر ادراری ساکن (IUT) اشاره کرد. در این مطالعه، ترکیب CAI قبلی و HAI یک متغیر منفرد (یعنی عفونت‌های قبلی) در نظر گرفته شده است. این اطلاعات به عنوان مجموعه ویژگی‌های اولیه در نظر گرفته شده‌اند. با استفاده از روش‌های آماری و همچنین تحلیل هم‌بستگی، ارتباط احتمالی بین ویژگی‌های اولیه بررسی شد و روش‌های کاهش بُعد داده مانند تجزیه و تحلیل مؤلفه‌های اصلی (PCA) و تجزیه و تحلیل مؤلفه‌های همسایگی (NCA) [۱۷] برای کاهش اندازه فضای ویژگی استفاده شدند.

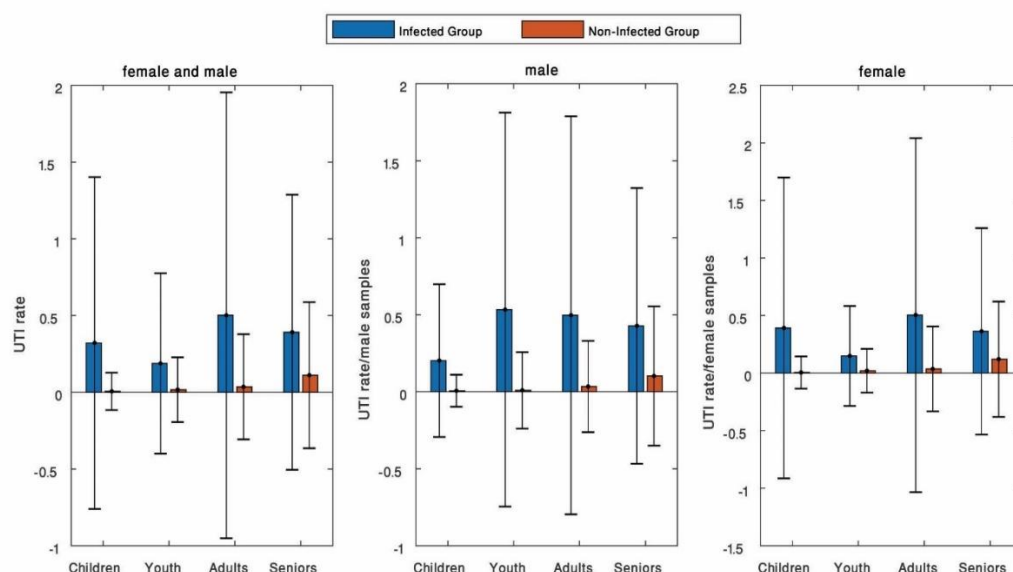
الگوریتم‌های یادگیری ماشین و طبقه‌بندی

هدف اصلی مطالعه حاضر ایجاد یک مدل پیش‌بینی براساس اطلاعات دموگرافیک و سوابق بیماری بیماران بستری شده در بیمارستان جهت پیش‌بینی احتمال عفونت ادراری است. برای این منظور، انواع مختلفی از طبقه‌بندها با ناظر شامل درخت‌های تصمیم‌گیر (decision tree)، ماشین‌های بردار پشتیبان (SVM)، نزدیک‌ترین K همسایگی (KNN)، آنالیز تفکیک‌کننده خطی (discriminant analysis) و طبقه‌بندی رگرسیون خطی مورد استفاده قرار گرفتند. با استفاده از روش‌های بهینه‌سازی پارامترها که با انتخاب تصادفی ۱۰٪ از داده‌های آموزش و روش جست‌وجوی تصادفی صورت پذیرفته است، پارامترهای بهینه هر روش یادگیری ماشین یافت شده‌اند. براین اساس، برای طبقه‌بند درخت تصمیم‌گیر، حداقل اندازه برگ ۳ و حداکثر تعداد تقسیم‌بندی ۷ تنظیم شد. برای طبقه‌بندی‌کننده SVM، از یک

۵۸/۴۵ درصد موارد در گروه مبتلا زن و در گروه سالم ۵۳/۵۳ درصد زن بودند. به منظور بررسی همبستگی بین عوامل، تجزیه و تحلیل آماری و همچنین تحلیل همبستگی انجام شد. متغیرهای حشو Redundant Variables را می توان داده های زائد در نظر گرفت که تأثیر قابل توجهی بر عملکرد طبقه بندی نخواهند داشت. تجزیه و تحلیل آماری با استفاده از آزمون کای ۲ برای مقایسه نسبت برابر COPD (χ^2 برابر ۰/۷۵۲/۲۵ p کوچکتر از ۰/۰۰۱)، احتباس ادرار (χ^2 برابر ۱۶۵۶/۹ p کوچکتر از ۰/۰۰۱) و اختلالات عصبی (χ^2 برابر ۸۰۴ p کوچکتر از ۰/۰۰۱) فرضیه صفر را رد کرد. این نتیجه نشان دهنده اهمیت این عوامل مخدوش کننده برای تمایز بین UTI و موارد سالم (شاهد) بود. برای آزمون اثر متقابل احتمالی بین این سه عامل، از آزمون کوکران مانتل هانزل استفاده شد.

همچنین، تحلیل همبستگی بین زوج متغیرها با استفاده از همبستگی پیرسون انجام شد. نتایج همبستگی ضعیفی را بین این عوامل نشان داد [COPD در مقابل احتباس ادرار (مقدار همبستگی پیرسون برابر ۰/۰۳۵ p کوچکتر از ۰/۰۰۱)؛ COPD در مقابل

اختلالات عصبی (مقدار همبستگی برابر ۰/۰۵۳ p کوچکتر از ۰/۰۰۱)؛ احتباس ادرار در مقابل اختلالات عصبی (مقدار همبستگی برابر ۰/۰۴۷ p کوچکتر از ۰/۰۰۱)]. این آزمون ها نشان داد که بین سه عامل COPD، احتباس ادرار و اختلالات عصبی اثر متقابل بسیار کوچک (کوچکتر از ۰/۵)، اما معناداری وجود دارد. در این راستا، این متغیرها را می توان متغیرهای غیرهمبسته ای در نظر گرفت که آن ها را برای اهداف طبقه بندی مناسب می کند. علاوه بر این، تجزیه و تحلیل همبستگی نشان داد که هیچ رابطه خطی بین عفونت قبل از بستری، حین بستری و موارد کاتتر ادراری ساکن وجود ندارد (مقدار همبستگی برابر صفر). با در نظر گرفتن سن نمونه ها، چهار گروه کودکان (۰-۱۴ سال)، جوانان (۱۵-۲۴ سال)، بزرگسالان (۲۵-۶۴ سال) و سالمندان (بیش از ۶۵ سال) در نظر گرفته شدند. آزمون t مستقل نشان داد که تفاوت معنی داری در تعداد دفعات بروز عفونت های ادراری پیشین بین نمونه های مبتلا و غیرمبتلا در بازه های سنی مختلف وجود ندارد ($p < 0/01$). مقایسه گرافیکی میزان UTI در شکل ۱ برای جنسیت ها و بازه های سنی مختلف نشان داده شده است.



شکل ۱- مقایسه میانگین تعداد عفونت های پیشین UTI بین بازه های سنی مختلف برای گروه های مبتلا و سالم

نتایج طبقه بندی

با در نظر گرفتن مجموعه ویژگی های اصلی (سن، جنس، بستری مجدد، COPD، احتباس ادرار، بیماری عصبی، تعداد UTI های قبلی و تعداد IUC)، نتایج طبقه بندی با استفاده از انواع مختلف طبقه بندی ها برای پیش بینی UTI در جدول ۱ گزارش شده است. براساس نتایج جدول ۱، حداکثر دقت به طبقه بندی درخت تصمیم گیر مربوط می شود؛ به نحوی که با کاهش ابعاد ویژگی بارش NCA و PCA دقت به ترتیب ۷۸/۷۱٪ و ۵۶/۶۹٪ بوده است. دقت بالاتر طبقه بندی درخت تصمیم گیر برای تکرارهای مختلف

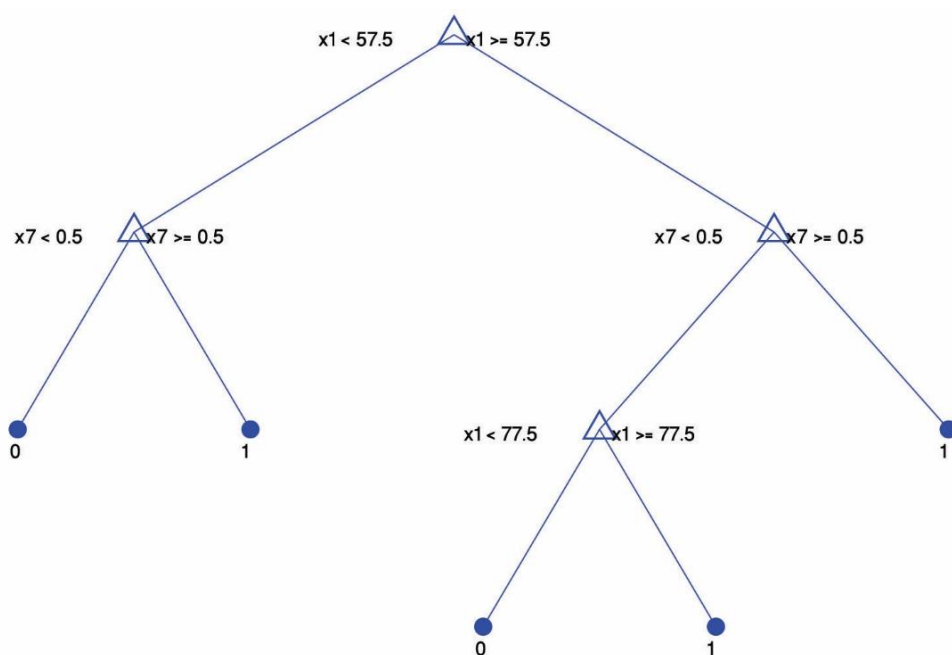
طبقه بندی (با استفاده از مجموعه های مختلف داده های آموزش و تست) همراه با تغییرپذیری نسبتاً بالایی است، درحالی که سایر طبقه بندی ها نتایجی با تغییرپذیری کمتر به دست می دهند. طبق نتایج جدول ۱، طبقه بندی بعدی با دقت بالاتر، طبقه بندی تفکیک کننده خطی بود. در این راستا، این دو نوع طبقه بندی (درخت تصمیم گیر و تفکیک کننده خطی) برای تفسیر نتایج انتخاب شدند. از آنجایی که حساسیت (نرخ تشخیص درست موارد مبتلا به عفونت ادراری) بیشتر از ویژگی (نرخ تشخیص درست موارد غیرمبتلا به عفونت ادراری) مدنظر بوده است، طبقه بندی تفکیک کننده خطی با مجموعه

خطر بیشتری برای UTI بودند. درحالی که برای افراد کمتر از ۵۷/۵ سال و خطر کمتر UTI قبلی، احتمال UTI ممکن است به طور قابل توجهی کاهش یابد. برای افراد بالای ۵۷/۵ سال، سن بحرانی دیگر ۷۷/۵ سال بود که در آن خطر عفونت ادراری بدون سابقه قبلی UTI نیز قابل توجه بوده است.

ویژگی اولیه (بدون کاهش بُعد) در این مطالعه انتخاب شد (جدول ۱ را ببینید). با توجه به شکل ۲، دو متغیر سن (X_1) و تعداد UTI های قبلی (X_7) عوامل خطر مهمی بودند که پیش بینی UTI های آینده را امکان پذیر می کنند. سن ۵۷/۵ سال یک مقدار بحرانی بود، به گونه ای که افراد بالای ۵۷/۵ سال و عفونت های قبلی در معرض

جدول ۱. عملکرد طبقه بندی های مختلف برای پیش بینی UTI براساس داده های دموگرافیک و سابقه بیماری؛ خطای طبقه بندی با استفاده از رهیافت ۱۰- fold cross-validation در فاز آموزش و به صورت (۱-دقت) $10 \times$ محاسبه شده است. برای این منظور، داده های آموزش با نسبت ۸۰ به ۲۰ جهت آموزش و اعتبارسنجی به صورت تصادفی تفکیک شده اند.

ویژگی	طبقه بندی	خطای طبقه بندی برای داده های اعتبارسنجی	حساسیت	ویژگی	دقت	AUC Area Under (Curve)
Original feature set	Decision Tree	۲۹/۰۴±۰/۱۸	۷۴/۹۲±۱۳/۰۶	۷۴/۹۲±۱۳/۰۶	۶۷/۷۶±۱۰/۶۶	۰/۷۱±۰/۰۰
	SVM	۲۸/۱۶±۰/۰۸	۶۱/۵۰±۱/۳۲	۸۲/۲۰±۱/۵۲	۶۳/۵۷±۱/۰۴	۰/۷۲±۰/۰۰
	KNN	۲۹/۹۱±۰/۲۹	۵۴/۶۰±۰/۷۰	۸۵/۹۹±۰/۸۹	۵۷/۷۴±۰/۶۱	۰/۷۰±۰/۰۰
	Discriminant analysis	۲۷/۹۹±۰/۱۴	۸۰/۹۳±۱/۰۲	۶۳/۴۱±۰/۷۹	۶۵/۱۶±۰/۶۴	۰/۷۲±۰/۰۱
	Linear classifier	۲۸/۲۶±۰/۵۹	۶۱/۶۶±۰/۶۴	۸۲/۱۴±۱/۰۹	۶۳/۷۱±۰/۵۵	۰/۷۲±۰/۰۰
Reduced feature set by PCA	Decision Tree	۲۹/۱۳±۰/۲۷	۶۹/۳۰±۱۳/۱۵	۷۱/۹۱±۱۲/۷۸	۶۹/۵۶±۱۰/۵۷	۰/۷۱±۰/۰۰
	SVM	۲۸/۱۷±۰/۰۹	۶۱/۳۸±۱/۷۹	۸۱/۹۲±۲/۲۹	۶۳/۴۳±۱/۴۲	۰/۷۲±۰/۰۰
	KNN	۲۹/۵۹±۰/۲۶	۵۵/۰۷±۰/۶۶	۸۵/۳۱±۱/۰۲	۵۸/۰۹±۰/۵۹	۰/۷۰±۰/۰۰
	Discriminant analysis	۲۷/۹۷±۰/۱۶	۶۳/۵۲±۰/۷۹	۸۰/۸۳±۱/۰۳	۶۵/۲۵±۰/۶۷	۰/۷۲±۰/۰۰
	Linear classifier	۲۸/۲۹±۰/۶۵	۵۶/۷۷±۱۷/۵۸	۸۳/۱۵±۶/۵۸	۵۹/۴۱±۱۵/۲۱	۰/۷۰±۰/۰۶
Reduced feature set by NCA	Decision Tree	۲۹/۲۲±۰/۲۲	۷۲/۰۱±۱۲/۰۲	۶۹/۷۴±۱۱/۷۴	۷۱/۷۸±۹/۶۵	۰/۷۱±۰/۰۰
	SVM	۲۸/۶۲±۰/۰۰	۵۴/۷۹±۰/۴۱	۸۶/۶۳±۰/۴۲	۵۷/۹۷±۰/۳۷	۰/۷۱±۰/۰۰
	KNN	۲۹/۷۲±۰/۳۲	۵۴/۸۲±۰/۷۱	۸۵/۹۴±۱/۰۷	۵۷/۹۳±۰/۶۰	۰/۷۰±۰/۰۰
	Discriminant analysis	۲۸/۰۲±۰/۱۲	۶۳/۵۴±۰/۳۶	۸۰/۸۶±۰/۴۲	۶۵/۲۷±۰/۲۹	۰/۷۲±۰/۰۰
	Linear classifier	۲۸/۱۲±۰/۲۶	۶۱/۲۵±۰/۴۵	۸۲/۳۱±۰/۵۹	۶۳/۳۶±۰/۴۳	۰/۷۲±۰/۰۰



شکل ۲. شماتیک نمودار طبقه بندی درخت تصمیم گیر با بهترین دقت طبقه بندی: X_1 سن، و X_7 تعداد UTI قبلی

ویژگی ها با توجه به اهمیت آن ویژگی در طبقه بندی به دست

روش NCA برای کاهش بُعد ویژگی، وزنی را برای هریک از

می‌آورد. وزن‌های به‌دست‌آمده در جدول ۲ گزارش شده‌اند. براساس جدول ۲، بااهمیت‌ترین ویژگی‌ها براساس NCA به ترتیب سن، سابقه قبلی عفونت ادراری، بستری مجدد و بیماری عصبی بودند. سایر ویژگی‌ها مانند جنسیت، COPD، احتباس ادرار و IUC های قبلی سهم قابل توجهی در طبقه‌بندی نداشتند.

نتایج شکل ۲ و جدول ۲ تأثیر قابل توجه سن را بر پیش‌بینی UTI نشان می‌دهند. به‌عنوان تحلیل دیگری، سن از فضای ویژگی حذف شد و طبقه‌بندی مجدداً انجام گرفت (طبقه‌بندی: درخت تصمیم‌گیر؛ روش کاهش بُعد فضای ویژگی: NCA). تنها ویژگی‌هایی که با روش NCA وزنی بالاتر از ۰/۱ را به دست داده‌اند، برای طبقه‌بندی استفاده شدند. نتایج نشان داد که حذف سن از فضای ویژگی حساسیت و دقت طبقه‌بندی را به ترتیب به ۹۱/۲۰٪ (۰/۲۱) و ۸۵/۲۰٪ (۰/۱۸) افزایش داد، درحالی‌که ویژگی به‌طور قابل توجهی به ۳۲/۴۵ (۰/۹۲) درصد و سطح زیر ROC به ۰/۶۲ کاهش یافت. در چنین مواردی، UTI قبلی، IUC قبلی، بیماری عصبی و احتباس ادراری بااهمیت‌ترین فاکتورها در پیش‌بینی عفونت ادراری بودند (به ترتیب دارای وزن ۲/۸۵، ۲/۵۰، ۱/۹۸ و ۱/۸۹). جنسیت و IUC قبلی مجدداً برای تشخیص اهمیت نداشتند (وزن برابر ۰). علاوه‌براین، هنگامی که از طبقه‌بند تفکیک‌کننده خطی (بدون

کاهش ویژگی) استفاده شد، دقت و حساسیت به ترتیب ۸۳/۰±۹۱/۲۷ و ۸۹/۳۸±۰/۲۵ درصد بود، درحالی‌که ویژگی به ۳۴/۰±۷۱/۸۹ درصد کاهش یافت. این تجزیه‌وتحلیل‌ها همراه با نتایج شکل ۲ ممکن است نشان دهد که سن شاید باید به‌عنوان یک عامل تمایزدهنده در نظر گرفته شود. دراین‌راستا، طبقه‌بندی مجدد برای سنین مختلف [کودکی (۰-۱۴)، جوانی (۱۵-۲۴)، بزرگ‌سالی (۲۵-۶۴) و سالمندی (۶۵ و بالاتر)] انجام شد و نتایج در جدول ۳ گزارش شده است.

جدول ۲. وزن ویژگی‌ها در مدل پیش‌بینی UTI براساس روش NCA

ویژگی	وزن
Age	۵/۲۹
Gender	۰
Readmission	۱/۷۲
COPD	۰
Urinary retention	۰
Neurological disease	۱/۰۷
Previous UTI	۲/۷۵
Previous IUC	۰

جدول ۳. عملکرد طبقه‌بندی UTI برای زیرگروه‌های سنی مختلف؛ بااهمیت‌ترین ویژگی‌ها برای هر زیرگروه در ستون آخر گزارش شده است.

ویژگی‌های مهم	ROC	ویژگی	حساسیت	دقت	زیرگروه (%)
Previous UTI; Gender; urinary retention; Readmission	۰/۵۹±۰/۰۶	۵۹/۰۰±۲۵/۵۸	۶۰/۷۸±۲۱/۹۴	۶۰/۶۰±۱۷/۴۷	کودک ۱۳/۰۷
Previous UTI; Gender; Readmission; previous IUC; neurological disease	۰/۶۳±۰/۰۳	۹۳/۳۳±۴/۷۱	۳۳/۲۴±۲/۵۹	۳۹/۴۲±۲/۴۱	جوان ۶/۱۴
Previous UTI; previous IUC; Gender; neurological disease; COPD	۰/۶۳±۰/۰۲	۳۴/۸۴±۳/۱۴	۹۱/۹۷±۰/۳۰	۸۶/۲۵±۰/۳۶	بزرگ‌سال ۳۹/۱۲
Previous UTI; previous IUC; Readmission;	۰/۵۹±۰/۰۰	۴۰/۸۹±۱۶/۶۹	۷۷/۲۷±۱۶/۵۵	۷۳/۶۴±۱۳/۲۳	کهن‌سال ۴۱/۶۷

نتایج

مهم‌ترین نتایج مطالعه حاضر اهمیت سن در پیش‌بینی عفونت ادراری فارغ از جنسیت و همچنین توانایی الگوریتم‌های یادگیری ماشین در پیش‌بینی عفونت ادراری با تکیه بر تعداد محدودی ویژگی بوده است. تجزیه‌وتحلیل‌های آماری نشان داد که تعداد قبلی عفونت‌های ادراری بین موارد فعلی UTI و موارد غیر UTI تفاوت معنی‌داری نداشت. باین‌حال، همان‌طور که شکل ۱ نشان می‌دهد، میانگین تعداد UTI های قبلی در افراد مبتلای فعلی در مقایسه با نمونه‌های غیرمبتلای فعلی بیشتر بود. این مسئله نشان می‌دهد که سابقه عفونت باید به‌عنوان یک عامل مهم برای پیش‌بینی UTI در نظر گرفته شود و این پیامد تحت تأثیر جنسیت نیز نخواهد بود.

تعداد موارد ابتلا به عفونت برای هر دو گروه «مبتلا» و «غیرمبتلا» عمدتاً در بزرگ‌سالی (بزرگ‌سالان و سالمندان) در مقایسه با نمونه‌های جوان‌تر (کودکان و افراد جوان) بیشتر بود. برای نمونه‌های غیرمبتلا، میزان بروز نسبت به سن افزایش نشان می‌داد، درحالی‌که برای نمونه‌های مبتلا یک رابطه غیرخطی بین سن و میزان ابتلا مشاهده شد. برخی از مطالعات کاهش بروز UTI را در سنین میان‌سالی و سپس افزایش میزان آن با افزایش سن را گزارش کرده‌اند [۱۹].

مطابق جدول ۱، دقت پیش‌بینی UTI بین ۵۷ تا ۷۱ درصد متغیر بود. این مسئله نشان می‌دهد که ویژگی‌های ساده دموگرافیک و سابقه بیماری بهتر از یک تصمیم تصادفی می‌توانند

ترکیبی از الگوریتم ژنتیک و شبکه عصبی مصنوعی برای پیش‌بینی UTI استفاده کردند. حجم نمونه برای مطالعه مذکور ۲۱۲ زن در محدوده سنی ۱۹ تا ۸۴ سال بود. متغیرهای پیش‌بینی‌کننده برای این مطالعه تکرر ادرار، آزمایش خون و فاکتورهای استخراج‌شده از آزمایش ادرار بودند. دقت به‌دست‌آمده ۷۶/۴٪ با فاصله اطمینان ۹۵٪ برابر با ۷۰-۸۱/۸ بود [۲۲]. مولر و همکاران (۲۰۲۱) با مجموعه ویژگی‌های مشابه مطالعه ما، طبقه‌بندی با ROC از ۰/۸۲ تا ۰/۸۴ برای موارد مبتلا به UTI در جامعه و ۰/۷۱ تا ۰/۷۷ برای موارد بیمارستانی مبتلا به UTI پیشنهاد کردند [۱۴]. این مقدار برای مطالعه حاضر ۰/۷ تا ۰/۷۲ بوده است.

محدودیت‌های مطالعه

داده مورد استفاده در این مطالعه دربردارنده تعداد بسیار محدودی ویژگی بوده است. این مسئله بر روی دقت طبقه‌بند جهت پیش‌بینی عفونت ادراری تأثیرگذار بوده است. ازسوی دیگر، با توجه به نامتعادل بودن گروه‌ها در داده مورد استفاده که می‌تواند سبب سوگیری نتایج شود، از روش‌های کاهش داده استفاده شد که خود می‌تواند سبب از دست رفتن بخشی از اطلاعات موجود در داده‌گان شود.

نتیجه‌گیری

بیماری‌های عفونت ادراری از شایع‌ترین عفونت‌ها در جوامع بشری هستند. استاندارد طلایی برای تشخیص چنین عفونت‌هایی، آزمایش کشت ادرار است که فرایندی زمان‌بر و مستعد خطا خواهد بود. هر تحلیلی که به پیش‌بینی، تشخیص یا یافتن موارد مشکوک کمک کند، ممکن است هزینه کار تشخیصی را کاهش و دقت نتیجه را افزایش دهد. در این مطالعه، پتانسیل اطلاعات دموگرافیک و سوابق پزشکی بیمار مشکوک به عفونت ادراری برای پیش‌بینی احتمال عفونت ادراری بررسی شد. این دسته ویژگی‌ها به‌سهولت در دسترس خواهند بود. نتایج این مطالعه نشان داد که چنین ویژگی‌های ساده‌ای می‌توانند سیستم‌های هوشمندی را برای پیش‌بینی احتمال ابتلا به UTI برای یک بیمار بستری در بیمارستان ارائه دهند. حتی اگر دقت پیش‌بینی خیلی بالا نباشد (تا ۷۱٪ در مطالعه حاضر)، نتیجه مطالعه حاضر بر برتری سیستم‌های هوشمند بر پیش‌بینی تصادفی UTI تأکید دارد. چنین سیستم هوشمندی را می‌توان با ویژگی‌های پیچیده‌تر برای افزایش دقت، حساسیت و ویژگی پیش‌بینی غنی کرد.

تشکر و قدردانی

از معاونت تحقیقات و فناوری دانشگاه علوم پزشکی همدان کمال تشکر را جهت پشتیبانی به‌عمل‌آمده داریم. مطالعه حاضر مستخرج از طرح تحقیقاتی ۱۴۰۳۰۸۱۵۷۰۰۷ است.

تضاد منافع

هیچ‌گونه تضاد منافی در پژوهش حاضر وجود ندارد.

در پیش‌بینی UTI عمل کنند. برای انواع مختلف مجموعه ویژگی‌ها و طبقه‌بندها، مشخصه ویژگی طبقه‌بند عمدتاً بهتر از حساسیت آن بود. این بدان معناست که طبقه‌بند بهتر می‌تواند موارد منفی واقعی (پیش‌بینی درست موارد غیرمبتلا) را در مقایسه با موارد مثبت درست (پیش‌بینی واقعی موارد مبتلا) پیش‌بینی کند. با مقایسه دقت طبقه‌بند درخت تصمیم‌گیر (گزینه با حداکثر دقت)، زمانی که کاهش بُعد فضای ویژگی با استفاده از روش‌های PCA یا NCA اعمال شد، عملکرد طبقه‌بند می‌تواند بهبود یابد. برای سایر طبقه‌بندها، کاهش ویژگی تأثیر معنی‌داری را بر روی بهره‌وری تشخیص نشان نداد. این مسئله ممکن است به دلیل رابطه مستقل بین متغیرهای در نظر گرفته‌شده باشد.

با توجه به شکل ۲، دو سن بحرانی برای پیش‌بینی UTI (۵۷/۵ و ۷۷/۵ سال) برای داده مورد تجزیه و تحلیل در این مطالعه به دست آمده است؛ به‌گونه‌ای که عفونت UTI پس از این سنین به‌طور قابل‌توجهی افزایش نشان می‌دهد. مطالعات قبلی گزارش کرده‌اند که میزان بروز برای ۶۵-۷۴ سال ۰/۰۵ در هر فرد در سال افزایش و برای مردان و زنان، بروز UTI در سنین بالای ۸۵ سال به‌طور قابل‌توجهی جهش می‌یابد [۲۰].

هنگامی که تمام بازه‌های سنی در نظر گرفته شدند (کودکی و بزرگ‌سالی)، ویژگی جنسیت تأثیر ناچیزی بر پیش‌بینی UTI داشت (جدول ۲ را ببینید). درحالی‌که برای دوره‌های سنی مختلف، جنسیت باید به‌عنوان یک ویژگی مهم در نظر گرفته شود (جدول ۳ را ببینید). با توجه به متون، UTI در زنان شایع‌تر است. اما در بین افراد مسن تفاوتی بین جنس‌های مختلف در ابتلا به UTI مشاهده نمی‌شود [۲۱]. این مسئله مطابق با نتایج گزارش‌شده در جدول ۳ بود که در آن جنسیت به‌عنوان یک ویژگی دارای اهمیت برای سنین مختلف، به‌جز برای سالمندان، بود. علاوه‌براین، طبق نتایج جدول ۳، دقت طبقه‌بند پیش‌بینی‌کننده UTI برای گروه‌های بزرگ‌سال و سالمند در مقایسه با گروه‌های کودک و جوان به‌طور معنی‌داری بیشتر بود. این نتایج و وزن بااهمیت‌ترین ویژگی‌ها برای دوره‌های سنی مختلف (جدول ۳) تأکید می‌کند که هر سیستم هوشمند برای پیش‌بینی UTI باید تصمیم را به شیوه‌ای متفاوت برای دوره‌های سنی مختلف اتخاذ کند.

با توجه به جدول ۱، بهترین دقت طبقه‌بندی توسط طبقه‌بند درخت تصمیم‌گیر (۷۱/۷۸) برای مجموعه ویژگی‌های کاهش‌یافته به دست آمد. درحالی‌که یک طبقه‌بند تفکیک‌کننده خطی بدون هیچ کاهشی در بُعد مجموعه ویژگی‌های اولیه، دقت ۶۵/۱۶ را به دست داد. در مقایسه با عملکرد سیستم پیش‌بینی‌کننده پیشنهادی در این مطالعه، اوزکان و همکاران (۲۰۱۸) از داده‌های معاینه معمولی (شامل سوابق بیماری، معاینه بالینی و داده‌های معاینه اولتراسوند) استفاده کردند و با استفاده از شبکه‌های عصبی مصنوعی، دقت طبقه‌بندی بالای ۹۸/۳۰ درصد را به دست آوردند [۱۰]. با این حال، حجم نمونه مطالعه مذکور نسبتاً پایین و به اندازه ۵۹ نفر بوده است. برای زنان، هکرلینگ و همکاران (۲۰۰۷) از

تجزیه و تحلیل داده و نگارش نسخه اولیه مقاله را سجاد فراشی و تحلیل نتایج را حسین عماد ممتاز برعهده داشته‌اند.

حمایت مالی

حمایت مالی این مطالعه از جانب دانشگاه علوم پزشکی همدان به انجام رسیده است (شماره طرح: ۱۴۰۳۰۸۱۵۷۰۰۷).

ملاحظات اخلاقی

این مطالعه با کد اخلاق ۵۶۹، ۱۴۰۳ IR.UMSHA.REC. در کمیته اخلاق دانشگاه علوم پزشکی همدان ثبت شده است.

REFERENCES

- Barnett BJ, Stephens DS. Urinary tract infection: an overview. *Am J Med Sci*. 1997;314(4):245-249. PMID: 9332263 DOI: 10.1097/00000441-199710000-00007
- Goździkiewicz N, Zwolińska D, Polak-Jonkisz D, Zhang F, Mo... The use of artificial intelligence algorithms in the diagnosis of urinary tract infections-a literature review. *J Clin Med*. 2022;11(10):2734. PMID: 35930733 DOI: 10.3390/jcm11102734
- Schmiemann G, Kniehl E, Gebhardt K, Matejczyk MM, Hummers-Pradier E. The diagnosis of urinary tract infection: a systematic review. *Dtsch Arztebl Int*. 2010;107(21):361-367. PMID: 20539810
- Storme O, Tirán Saucedo J, Garcia-Mora A, Dehesa-Dávila M, Naber KG. Risk factors and predisposing conditions for urinary tract infection. *Ther Adv Urol*. 2019;11:1756287218814382. PMID: 31105772 DOI: 10.1177/1756287218814382
- Nitzan O, Elias M, Chazan B, Saliba W. Urinary tract infections in patients with type 2 diabetes mellitus: review of prevalence, diagnosis, and management. *Diabetes Metab Syndr Obes*. 2015;8:129-136. PMID: 25759592 DOI: 10.2147/DMSO.S51792
- Scholes D, Hooton TM, Roberts PL, Stapleton AE, Gupta K, Stamm WE. Risk factors for recurrent urinary tract infection in young women. *J Infect Dis*. 2000;182(4):1177-1182. PMID: 10979915 DOI: 10.1086/315827
- Hu KK, Boyko EJ, Scholes D, Normand E, Chen CL, Grafton J, et al. Risk factors for urinary tract infections in postmenopausal women. *Arch Intern Med*. 2004;164(9):989-993. PMID: 15136308
- Hufschmidt A, Shabarin V, Rauer S, Zimmer T. Neurological symptoms accompanying urinary tract infections. *Eur Neurol*. 2010;63(3):180-183. PMID: 20197663 DOI: 10.1159/000289098
- Goździkiewicz N, Zwolińska D, Polak-Jonkisz D. The use of artificial intelligence algorithms in the diagnosis of urinary tract infections-a literature review. *J Clin Med*. 2022;11(10):2734. PMID: 35628861 DOI: 10.3390/jcm11102734
- Ozkan IA, Koklu M, Sert IU. Diagnosis of urinary tract infection based on artificial intelligence methods. *Comput Methods Programs Biomed*. 2018;166:51-59. PMID: 30415718 DOI: 10.1016/j.cmpb.2018.10.007
- Burton RJ, Albur M, Eberl M, Cuff SM. Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. *BMC Med Inform Decis Mak*. 2019;19. DOI:10.1186/s12911-019-0878-9
- Advanced Analytics Group of Pediatric Urology and ORC Personalized Medicine Group. Targeted workup after initial febrile urinary tract infection: using a novel machine learning model to identify children most likely to benefit from voiding cystourethrogram. *J Urol*. 2019;202(1):144-152. PMID: 30810465 DOI: 10.1097/JU.000000000000186
- Gadalla AAH, Friberg IM, Kift-Morgan A, Zhang J, Eberl M, Topley N, et al. Identification of clinical and urine biomarkers for uncomplicated urinary tract infection using machine learning algorithms. *Sci Rep*. 2019;9(1):19694. PMID: 31873085 DOI: 10.1038/s41598-019-55523-x
- Jeng SL, Huang ZJ, Yang DC, Teng CH, Wang MC. Machine learning to predict the development of recurrent urinary tract infection related to single uropathogen, *Escherichia coli*. *Sci Rep*. 2022;12(1):17216. PMID: 36241875
- Dong F, Yao Y, Chen Y, Guo Y, Jing C, Wu J. Diagnostic performance of urine analysis based on flow microimaging and artificial intelligence recognition technology in suspected urinary tract infection patients. *Scand J Clin Lab Invest*. 2022;82(5):385-390. PMID: 35852133 DOI: 10.1080/00365513.2022.2100273
- Møller JK, Sørensen M, Hardahl C. Prediction of risk of acquiring urinary tract infection during hospital stay based on machine-learning: A retrospective cohort study. *PLoS One*. 2021;16(3):e0248636. PMID: 33788888 DOI: 10.1371/journal.pone.0248636
- Yang W, Wang K, Zuo W. Neighborhood Component Feature Selection for High-Dimensional Data. *J Comput*. 2012;7(1): 161-168. DOI:10.4304/jcp.7.1.161-168
- MANTEL N, HAENSZEL W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719-748. PMID: 13655060 DOI:10.1093/inci/22.4.719
- Eriksson I, Gustafson Y, Fagerström L, Olofsson B. Prevalence and factors associated with urinary tract infections (UTIs) in very old women. *Arch Gerontol Geriatr*. 2010;50(2):132-135. PMID: 19349084 DOI: 10.1016/j.archger.2009.02.013
- Griebing TL. Urologic diseases in america project: trends in resource use for urinary tract infections in men. *J Urol*. 2005;173(4):1288-1294. PMID: 15758784 DOI: 10.1097/01.ju.0000155595.98120.8e
- Harrington RD, Hooton TM. Urinary tract infection risk factors and gender. *J Gend Specif Med*. 2000;3(8):27-34. PMID: 11253265
- Heckerling PS, Canaris GJ, Flach SD, Tape TG, Wigton RS, Gerber BS. Predictors of urinary tract infection based on artificial neural networks and genetic algorithms. *Int J Med Inform*. 2007;76(4):289-296. PMID: 16469531 DOI: 10.1016/j.ijmedinf.2006.01.005