مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره هفتم شماره دوم. ص ۱۶۰-۱۵۰، ۱۳۹۹

مقایسه الگوریتم‌های دسته‌بند در شناسایی چندداروییی و ویژگی‌های مؤثر بر آن در سالمندان

مرتضی مرادی[1]*، محمد مدرس[2]، محمدمهدی سپهری[3]

• دریافت مقاله: ۹۸/۵/۱۸ • پذیرش مقاله: ۹۸/۶/۲۵

مقدمه: تجویز و مصرف بیش از حد داروها که با عنوان چندداروییی شناخته می‌شود، هم موجب اتلاف منابع می‌گردد و هم برای بیماران زیان‌بار است. چندداروییی به‌خصوص برای سالمندان از اهمیت بیشتری برخوردار است؛ بنابراین عوامل مؤثر بر آن باید به‌درستی شناسایی و واکاوی شود.

روش: در این پژوهش گذشته‌نگر، نخست عملکرد الگوریتم‌های مختلف دسته‌بند C4.5، SVM، KNN، MLP و شبکه بیزی برای شناسایی چندداروییی، با نرم‌افزار WEKA مورد مقایسه قرار می‌گیرد. این فرآیند، با استخراج ۱۶ ویژگی جدید در کنار چهار ویژگی موجود در داده‌های ۸۱،۶۷۷ نسخه که برای تعداد ۱۹،۴۲۸ بیمار سرپایی با سن ۷۰ تا ۹۵ سال که در داروخانه‌های طرف قرارداد با بیمه سلامت استان تهران پیچیده شده‌اند، انجام شد. مقایسه عملکرد به‌وسیله آزمون t اصلاح شده با بازنمونه‌برداری صورت پذیرفت. به‌منظور شناسایی اثر ویژگی‌های بیماران بر چندداروییی، دو پارامتر مهم الگوریتم C4.5 به‌وسیله جستجوی توری بر روی ۵۰٪ مجموعه‌داده بهینه‌سازی و سپس بر ۵۰٪ دیگر مجموعه داده اعمال گردید و قوانین حاصل از آن در قالب درخت تصمیم و عبارات کلامی ارائه شد.

نتایج: مقایسه زوجی دسته‌بندها نشان‌گر عملکرد مناسب‌تر C4.5 و شبکه بیزی در مقایسه با سایر روش‌ها است. C4.5 توانایی شناسایی ویژگی‌های مؤثر بر چندداروییی را دارد. تنظیم پارامتر این الگوریتم باعث بهبود شاخص درستی و AUC شده و به‌شدت اندازه درخت تصمیم و تعداد قوانین تولیدی را کاهش می‌دهد.

نتیجه‌گیری: استفاده از رویکرد داده‌کاوی و به‌کارگیری C4.5 توانایی شناسایی و تبیین ویژگی‌های سالمندان را بر پدیده چندداروییی دارد. درصد مراجعه بیشتر به پزشکان عمومی و ارتباط با تعداد محدودتری از داروخانه از مهم‌ترین این ویژگی‌ها است.

کلید واژه‌ها: داده‌کاوی، دسته‌بندی، چندداروییی، منطقی‌سازی تجویز و مصرف دارو، سالمندان

•**ارجاع:** مرادی مرتضی، مدرس محمد، سپهری محمدمهدی. مقایسه الگوریتم‌های دسته‌بند در شناسایی چندداروییی و ویژگی‌های مؤثر بر آن در سالمندان مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۹؛ ۷(۲): ۶۰-۱۵۰.

۱. دانشجوی دکتری مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه صنعتی شریف، تهران، ایران
۲. دکتری مهندسی سیستم‌ها و تحقیق در عملیات، استاد، دانشکده مهندسی صنایع، دانشگاه صنعتی شریف، تهران، ایران
۳. دکتری تحقیق در عملیات، استاد، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

*** نویسنده مسئول:** مرتضی مرادی
آدرس: تهران، خیابان آزادی، دانشگاه صنعتی شریف، دانشکده مهندسی صنایع
•**شماره تماس:** ۰۲۱-۶۶۱۶۵۷۱۹
• **Email:** m.moradi@ie.sharif.edu

[Downloaded from jhbmi.com on 2026-11-17]

مقدمه

تجویز و مصرف هم‌زمان و بیش‌ازحد داروهای متعدد که با عنوان چنددارویی نیز شناخته می‌شود، هم موجب اتلاف منابع می‌شود و هم برای بیماران زیان‌بار است [۱،۲]. واکنش‌های نامطلوب دارویی (عوارض جانبی)، تداخل‌ها شامل تداخل دارو با دارو، دارو با مواد مغذی و دارو با سایر بیماری‌ها، همچنین عدم پذیرش دارودرمانی و عدم پیروی از دستورات دارویی از سوی بیمار از جمله عواقب چنددارویی است [۳]. Hovstadius و Petersson سن، جنسیت، نژاد/قومیت، موقعیت اقتصادی-اجتماعی، شرایط بالینی، شرایط دارودرمانی و ویژگی‌های رفتاری را به‌عنوان عوامل تأثیرگذار بیمار بر چنددارویی مطرح کرده است [۳]. چنددارویی و عوارض آن به‌خصوص برای بیماران سالمند از اهمیت بیشتری برخوردار است؛ زیرا به دلیل کهولت سن، بیشتر در معرض بیماری‌های مختلف و داروهای مرتبط با درمان آن‌ها قرار دارند. همچنین به دلیل تغییرات فیزیولوژیک مربوط به سالمندی، دارودرمانی برای آن‌ها با چالش‌های بیشتری همراه است [۴]؛ لذا بررسی چنددارویی و عوامل مؤثر بر آن، مورد توجه پژوهشگران زیادی قرار گرفته است. یافته‌های بررسی Galazzi و همکاران در کشور ایتالیا نشان داد که از ۲۰ تا ۶۵ درصد سالمندان دچار چنددارویی بوده‌اند و در معرض حداقل یک داروی بالقوه نامناسب قرار داشته‌اند [۵]. نتایج یک مطالعه مقطعی گذشته‌نگر در کشور عربستان که به بررسی میزان شیوع چنددارویی در بزرگسالان و سالمندان و عوامل مؤثر بر آن با استفاده از رگرسیون لجستیک چندمتغیره پرداخته، نشان‌گر آن است که میزان مصرف دارو در افراد دارای فشارخون و یا دارای دیابت دو برابر افراد بدون این بیماری‌ها است. در این پژوهش چنددارویی تحت تأثیر سن، جنس، وضعیت ازدواج و ملیت بیمار و وجود بیماری‌های چندگانه قرار داشت [۶]. Morin و همکاران نیز در یک مطالعه گروهی آینده‌نگر به همه‌گیرشناسی چنددارویی در حدود ۱،۷ میلیون نفر از افراد بالای ۶۵ سال در کشور سوئد در طول سه سال (از سال ۲۰۱۰ تا ۲۰۱۳) پرداخته‌اند. بر اساس تحقیقات آنان ۴۴ درصد افراد، در دوره مطالعه، به‌طور هم‌زمان بیشتر از ۵ دارو و ۱۱ درصد آن‌ها بیش از ۱۰ دارو مصرف می‌کرده‌اند [۷]. در پژوهش دیگری، برای تشخیص عوامل مؤثر بر چنددارویی شدید (مصرف هم‌زمان ۱۰ دارو و بیشتر) یک مطالعه مقطعی بر روی ۳،۴۰۹ بیمار سالمند انجام شده که نشانگر تأثیر عواملی همچون ضعف، تشخیص‌های چندگانه (بیش از ۸ تشخیص مختلف)، سن و شاخص توده‌بندی بر این چنددارویی شدید بود. از سوی دیگر جنسیت، سطح تحصیلات و مصرف سیگار بر آن مؤثر تشخیص داده نشدند [۸]. دلشاد و همکاران به بررسی ۳۲۳ نفر از سالمندان شهرستان گناباد به روش نمونه‌برداری طبقه‌ای تصادفی و با استفاده از پرسشنامه سه بخشی شامل ویژگی‌های دموگرافیک، الگوی مصرف دارو و عوامل مؤثر بر الگوی مصرف دارو پرداختند. یافته‌های آنان نشان داد که میانگین تعداد مصرف دارو ۴ قلم در شبانه‌روز است و ارتباط معنی‌دار بین چنددارویی با جنسیت، شغل و پیروی از دستورات دارویی وجود دارد [۹]. دیانتی و همکاران اطلاعات ۳۶۰ نفر از سالمندان شهر کاشان را بر اساس متغیرهای جمعیتی سن، جنسیت، وضعیت تأهل و نیز وضعیت اقتصادی، میزان تحصیلات و سابقه بستری در بیمارستان و متغیرهای اصلی بر اساس معیارهای Beers تحلیل کردند. نتایج بررسی آن‌ها بیان‌گر بروز نسبتاً شدید چنددارویی در سالمندان است، به‌گونه‌ای که میانگین تعداد داروی مصرفی ۶،۲ قلم دارو محاسبه شد [۱۰]. ترابی و همکاران نیز وضعیت تجویز داروهای نامناسب و چنددارویی در ۱۸۶ سالمند بالای ۶۰ سال بستری در بیمارستانی در شهر سبزوار را بر اساس معیارهای Beers بررسی کردند. نتایج بررسی آن‌ها نشان داد که حدود ۶۶ درصد بیماران بیش از ۵ دارو به‌طور هم‌زمان مصرف کرده بودند. البته آن‌ها رابطه معناداری بین مصرف هم‌زمان تعداد زیادی از داروها با ویژگی‌های جمعیتی شامل جنسیت و سن مشاهده نکردند [۱۱].

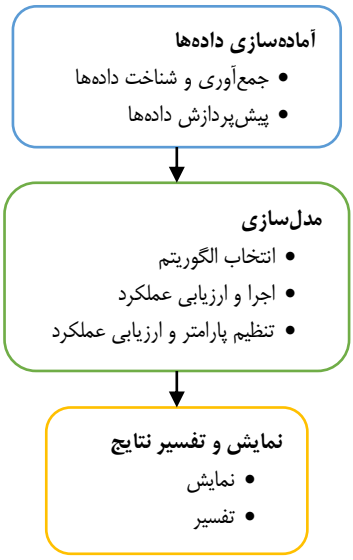
از سوی دیگر، بررسی پژوهش‌های انجام شده نشان‌گر آن است که علی‌رغم قابلیت‌های بسیار گسترده داده‌کاوی، این رویکرد در واکاوی چنددارویی و بررسی عوامل مؤثر بر آن تنها در تحقیقات معدودی مورد استفاده قرار گرفته است. در مطالعه‌ای، Held و همکاران با استفاده از قوانین همبستگی و تحلیل مجموعه‌های مکرر با هدف یافتن الگوهای جذاب از مصرف هم‌زمان داروها در میان ۱۶۸۶ بیمار مرد مسن‌تر از ۷۰ سال و ارتباط آن با سندروم سالمندی پرداختند [۱۲]. همچنین مرادی و همکاران از تکنیک‌های داده‌کاوی برای بررسی و تحلیل نسخه‌های تجویزی توسط متخصصان بیماری‌های داخلی [۱۳] و روانپزشکان [۱۴] به منظور شناسایی ویژگی‌های اثرگذار پزشکان در تجویز چنددارویی استفاده کردند.

[Downloaded from jhbmi.com on 2026-11-17]

با توجه به اهمیت موضوع و مسن‌تر شدن جمعیت کشور از یک سو و خلأ پژوهشی موجود در به‌کارگیری داده‌کاوی در چندارویی در سوی دیگر، در این مطالعه توانایی الگوریتم‌های دسته‌بند (به‌عنوان یکی از زیرمجموعه‌های مهم الگوریتم‌های داده‌کاوی) در شناسایی چندارویی و ویژگی‌های سالمندان که بر بروز چندارویی مؤثر است، بررسی شد.

روش

در این پژوهش گذشته‌نگر، پس از شناخت دقیق مسئله مورد بررسی و درک ابعاد مختلف آن، لازم است مراحل بعدی فرآیند داده‌کاوی به اجرا درآید. برای این منظور سه‌گام آماده‌سازی داده‌ها، مدل‌سازی و نیز نمایش و تفسیر نتایج انجام شد (شکل ۱).



شکل ۱: روش انجام پژوهش

آماده‌سازی داده‌ها: پیش از اجرای الگوریتم‌های یادگیری ماشین، لازم است تا آماده‌سازی داده‌ها انجام شود. این کار می‌تواند در دو مرحله صورت پذیرد. نخست، جمع‌آوری و شناخت داده‌ها انجام شد. به‌بیان‌دیگر، داده‌های مورد نیاز از منابع قابل‌اعتماد جمع‌آوری و با استفاده از تکنیک‌های پرس‌وجو مورد بررسی و شناخت دقیق قرار گرفتند. علاوه بر این، اگر داده‌ها از منابع اطلاعاتی مختلفی برداشت شده‌اند، لازم است تا با دقت کافی یکپارچه‌سازی شوند. در گام دوم، پیش‌پردازش داده‌های اولیه انجام شد و تلاش گردید تا با بهبود کیفیت داده‌ها شرایط لازم برای به‌کارگیری الگوریتم‌های مختلف کشف دانش فراهم شود. در پیش‌پردازش، داده‌های دارای نویز، ناموجود و یا ناسازگار شناسایی و اصلاح شدند. علاوه بر این، ویژگی‌های جدید مبتنی بر ویژگی‌های موجود، ایجاد شدند [۱۵،۱۶].

مدل‌سازی: در این مرحله بر اساس اهداف داده‌کاوی و متناسب با کیفیت و ویژگی‌های مجموعه داده‌های پیش‌پردازش شده از تکنیک‌ها و الگوریتم‌های مناسب داده‌کاوی برای اکتشاف و استخراج دانش پنهان موجود در داده‌ها استفاده شد.

در این مطالعه و با توجه به این‌که ویژگی «چندارویی» به عنوان یک ویژگی هدف برچسب‌دار تعریف می‌شود؛ لذا برای واکاوی داده‌ها از الگوریتم‌های دسته‌بندی استفاده شد. برای این منظور، ابتدا پنج دسته‌بند کارآمد و پر استفاده C4.5، SVM(Support Vector Machine)، K Nearest Neighbor (KNN، شبکه عصبی Multilayer Perceptron (MLP و Bayesian Network (BN انتخاب شده‌اند تا عملکرد هر یک از آن‌ها در شناسایی پدیده چندارویی مورد ارزیابی و مقایسه قرار گیرند. هر یک از این الگوریتم‌ها به گروهی متفاوتی از دسته‌بندها تعلق دارند که خصوصیات متفاوتی دارا می‌باشند. در واقع تلاش شده است که تأثیر خصوصیات متفاوت دسته‌بندها بر مجموعه داده‌های پژوهش مورد بررسی و تحلیل قرار گیرد تا بهترین الگوریتم در شناسایی چندارویی مشخص گردد.

برای اجرا و مقایسه این الگوریتم‌ها، از کاربرگ Experimenter در نرم‌افزار WEKA ویرایش ۳،۸،۲ استفاده شد. لازم به ذکر است که دسته‌بندهای مورد استفاده، با نام‌های متفاوتی در WEKA ارائه شده‌اند. به‌عبارت دقیق‌تر،

[Downloaded from jhbmi.com on 2026-11-17]

C4.5 با نام J48، SVM با عنوان LibSVM، KNN با نام IBK، MLP با عنوان MultilayerPerceptron و BN با عنوان BayesNet ارائه شده است. هر یک از این الگوریتم‌ها هم‌زمان با اعتبارسنجی ضربدری ۱۰ بخشی (10 Fold Cross-Validation) سه مرتبه بر روی مجموعه داده‌های مسئله اجرا شدند و سپس برای مقایسه عملکرد آن‌ها، آزمون زوجی t اصلاح شده با باز نمونه‌برداری در سطح معناداری ۹۵ درصد انجام شد. به‌بیان‌دیگر، کلیه الگوریتم‌ها بر اساس شاخص‌های عملکردی به صورت دو به دو با یکدیگر مقایسه شدند. شاخص‌های عملکردی مورد استفاده شامل درستی، حساسیت و شاخص AUC می‌شوند که بر اساس ماتریس آشفتگی تعریف می‌شوند.

معمولاً الگوریتم‌های مورد استفاده در داده‌کاوی، پارامترهای متعدد و قابل تنظیمی دارند که تغییر در مقادیر آن‌ها می‌تواند موجب تغییراتی در عملکرد الگوریتم‌ها شود. بر این اساس، تلاش برای تنظیم پارامترها در مقادیر بهینه توجیه‌پذیر است. روش‌های مختلفی برای تنظیم پارامتر توسعه داده شده‌اند. در این میان، جستجوی توری (شبکه‌ای) که مبتنی بر جستجوی کامل است یک روش مرسوم محسوب می‌شود. این روش با در نظر گرفتن ضرب کارتزین تعداد حالات ممکن برای مقادیر هر یک از پارامترهای مورد بررسی، الگوریتم مورد نظر را در هر یک از این حالت‌های مختلف اجرا می‌کند و نتایج حاصله را بر اساس یک سنجه از پیش تعیین شده با هم مقایسه کرده و بهترین ترکیب تنظیم پارامترها را مشخص می‌کند [۱۷]. از آنجا که فضای پارامتر (مقادیر مجازی که پارامتر می‌تواند به خود اختصاص دهد) برای برخی از پارامترهای خاص ممکن است شامل مقادیر حقیقی و یا نامحدود باشد، لازم است تا قبل از استفاده از الگوریتم جستجوی توری، برخی دست‌کاری‌ها و گسسته‌سازی‌هایی بر روی فضای پارامترها انجام شود. اگرچه در صورت وجود تعداد متعدد پارامترهای نیازمند به تنظیم، عملکرد جستجوی توری با مشکلاتی همراه است و به‌خصوص دچار مزاحمت ابعاد می‌شود؛ ولی در فضاهای پارامتری با ابعاد کم (مانند یک و یا دو بعد) عملکرد قابل اطمینانی ارائه می‌دهد [۱۸]. از آنجا که یکی از اهداف اصلی این پژوهش شناسایی عوامل مؤثر بروز چندارویی و شناخت خصوصیات و ویژگی‌های بیماران بر تمایل آن‌ها برای تجویز بیش از حد دارو است و از سوی دیگر، در میان الگوریتم‌های مورد بررسی تنها C4.5 قادر به انجام این موضوع است؛ بنابراین در ادامه دو پارامتر مهم و کلیدی آن با هدف بیشینه‌سازی شاخص درستی تنظیم شد. برای تنظیم پارامتر Confidence Factor مقادیر بین ۰/۰۰۱ تا ۰/۵۰۱ با طول گام ۰/۰۵ و برای تنظیم پارامتر minNumObj (که حداقل تعداد رکوردها در هر برگ از درخت تصمیم را مشخص می‌کند) مقادیر بین ۱۰۰ تا ۲۰۰ با طول گام ۵ مورد بررسی قرار گرفت. این بازه بدین دلیل در نظر گرفته شده تا از تولید درختان با ابعاد بزرگ جلوگیری شود و در نتیجه درخت کوچک‌تر و با قوانین ساده‌تر و خواناتری حاصل شود.

نمایش و تفسیر نتایج: به‌منظور آن‌که نتایج حاصل از داده‌کاوی برای کاربران و استفاده‌کنندگان نهایی پژوهش یعنی مدیران و تصمیم‌گیرندگان حوزه دارو، قابل فهم‌تر و قابل درک‌تر باشد، لازم است تا نتایج خروجی در قالب دقیق و مناسبی ارائه گردد. برای این منظور، نتایج داده‌کاوی که مجموعه قوانین استخراجی برای تبیین ویژگی‌های سالمندان در تمایل آنان به تجویزچندارویی است در قالب شکل و قوانین کلامی بیان شده و مورد تفسیر قرار خواهد گرفت.

نتایج

داده‌های اولیه مورد استفاده در این پژوهش از مرکز ملی تحقیقات بیمه سلامت و اداره کل بیمه سلامت استان تهران دریافت شد. این داده‌ها مربوط به نسخه‌های بیماران سرپایی بین ۷۰ تا ۹۵ سال سن است که از ابتدای سال ۱۳۹۴ تا انتهای آن، در داروخانه‌های طرف قرارداد با بیمه سلامت واقع در استان تهران پیچیده شده‌اند. به‌عبارت دقیق‌تر، اطلاعات عمومی مانند جنسیت، تاریخ تولد، نوع صندوق و نوع بیماری خاص بیماران از مرکز تحقیقات بیمه سلامت ایران دریافت شد. داده‌های نسخه‌های پزشکی نیز در قالب دو جدول در بانک اطلاعاتی MS SQL Server 2012 از اداره کل بیمه سلامت استان تهران گرفته شد. جدول نخست حاوی اطلاعات کلی نسخه مانند شماره نظام پزشکی، کد بیمار، تاریخ اعتبار دفترچه، تاریخ معاینه، تاریخ مراجعه به داروخانه و کد داروخانه است. جدول دوم نیز شامل مواردی مانند کد دارو، تعداد و قیمت داروهای موجود در هر نسخه است. اگرچه در جدول نخست فیلدهایی برای جنسیت بیمار و سن وی در نظر گرفته شده بود، ولی هیچ‌گونه اطلاعاتی در آن درج نشده بود؛ بنابراین تلاش شد تا اطلاعات تکمیلی بیماران اخذ شود. در نهایت فیلدهای اطلاعاتی جنسیت، تاریخ تولد، نوع صندوق بیمه، نوع بیماری خاص و محل صدور دفترچه برای تعدادی از بیماران

[Downloaded from jhbmi.com on 2026-11-17]

اخذ شد. همچنین داده‌های تکمیلی پزشکان، شامل نام و تخصص آنان نیز گردآوری شد. در ادامه کیفیت داده‌ها مورد ارزیابی قرار گرفت و رکوردهایی که دارای مقادیر نامعتبر و یا ناموجود بودند از مجموعه داده‌ها کنار گذاشته شدند. از سوی دیگر، از آنجا که ویژگی‌های موجود از کفایت لازم برای داده‌کاوی برخوردار نبودند، تعداد ۱۶ ویژگی‌های جدید بیماران از داده‌های موجود استخراج گردید. برای مثال، سن بیماران در سال ۱۳۹۴ و با توجه تاریخ تولد آن‌ها محاسبه شده و سپس بیماران با سن ۷۰ تا ۹۵ سال جداسازی شدند. همچنین با توجه به اسامی پزشکان و تطبیق آن با بانک اطلاعات اسامی، جنسیت آنان مشخص گردید و سپس درصد مراجعات هر بیمار به پزشکان زن یا مرد محاسبه شد. علاوه بر این، با توجه به آن‌که شماره نظام پزشکی عدد ترتیبی است و از عدد یک آغاز شده و به ترتیب به پزشکان متقاضی بعدی تخصیص داده شده است، می‌توان چنین گفت که پزشکان با شماره نظام پزشکی کوچک‌تر، زودتر به فعالیت پرداخته‌اند و با تجربه‌ترند. بر این اساس شماره نظام پزشکی در بازه‌هایی به طول ۴۰،۰۰۰ واحد گسسته‌سازی شد و سپس درصد مراجعه هر بیمار به هریک از این گروه‌ها محاسبه شد. همچنین تعداد نسخه‌های بیماران به انواع مختلف پزشکان بر حسب پزشک عمومی، پزشک متخصص (شامل متخصص، فلوشیپ و فوق تخصص) و نیز سایر پزشکان جداسازی شد و سپس درصد مراجعه هر بیمار به هریک از این دسته‌ها محاسبه گردید. به‌علاوه، به منظور بهبود کیفیت داده‌ها کاهش‌هایی در مقادیر برخی ویژگی‌ها به‌عمل آمد. برای مثال، محل صدور دفترچه بیماران از ۳۱ استان، به استان تهران و غیر تهران تنزل یافت و یا بیماران به دو دسته دارای بیماری خاص و عادی تفکیک شده و از نوع بیماری خاص به دلیل اندک بودن هریک از آن‌ها صرف‌نظر شد. در نهایت ویژگی چندارویی برای هر بیمار که در این پژوهش ویژگی برچسب دسته‌بندی است بدین‌صورت تعریف شد که درصورتی‌که میانگین تعداد داروهای تجویز شده در هر نسخه کمتر از حد آستانه سه [۲۰، ۱۹] قرار داشت این ویژگی عدد صفر را اختیار کرده و در غیر این صورت عدد یک به آن اختصاص داده شد. لازم به توضیح است که در خصوص حد آستانه برای چندارویی مقادیر مختلفی پیشنهاد شده است و وحدت نظر در این خصوص وجود ندارد [۲۱]. اطلاعات تکمیلی در خصوص ویژگی‌های بیماران که در مرحله مدل‌سازی مورد استفاده قرار گرفته است، در جدول ۱ نشان داده شده است.

در نهایت مجموعه داده‌ای با ۲۰ ویژگی برای تعداد ۱۹،۴۲۸ بیمار با سن ۷۰ تا ۹۵ سال برای داده‌کاوی آماده‌سازی شد که ۱۰،۵۹۳ نفر از آن‌ها (۵۴/۵٪) به‌طور متوسط سه دارو یا بیشتر از آن در هر نسخه برای آن‌ها تجویز شده است که نشان‌گر بروز چندارویی برای آن‌ها است. درحالی‌که ۸/۸۶۵ نفر (۴۵/۶٪) نیز به‌طور متوسط کمتر از سه دارو در هر نسخه داشته‌اند. همچنین تعداد ۱۰،۸۰۷ نفر بیمار زن و ۸،۶۵۱ نفر بیمار مرد در مجموعه داده‌ها وجود داشته‌اند. از طرفی، تنها تعداد ۱۵۹ نفر از آنان دارای بیماری خاص بوده‌اند. دفترچه بیمه سلامت ۱۴،۹۳۲ نفر در استان تهران و ۴/۵۲۶ نفر نیز در سایر استان‌ها صادر شده بود. لازم به توضیح است که این اطلاعات از بررسی و تحلیل ۸۱،۶۷۷ نسخه که برای این افراد پیچیده شده بود، به دست آمده است که نشان دهنده میانگین ۴,۲ نسخه برای هر بیمار در طول یک سال است.

در ادامه و با هدف کشف چندارویی، شش دسته‌بند مختلف بر روی مجموعه داده به اجرا درآمد و بر اساس شاخص عملکردی حساسیت، درستی و (Area Under the ROC Curve) AUC به صورت دو به دو با هم مقایسه شدند. این مقایسه به‌وسیله WEKA Experimenter و با در نظر گرفتن مقادیر پیش فرض برای کلیه پارامترهای الگوریتم‌ها انجام شده است. هر یک از این دسته‌بندها هم‌زمان با اعتبارسنجی ضربدری ۱۰ بخشی، سه مرتبه بر روی مجموعه داده‌های داده‌ها اجرا شده‌اند و سپس آزمون زوجی t اصلاح شده در سطح معناداری ۹۵ درصد انجام شده است. نتایج این بررسی در جداول ۲ تا ۴ نشان داده شده است. در این جداول اعداد بیرون پرانتز متوسط مقدار هر شاخص است، در حالی که اعداد داخل پرانتز بیان‌گر انحراف معیار آن است. همچنین حرف b نشان‌گر آن است که دسته‌بندی که در سطر قرار دارد نسبت به دسته‌بندی که در ستون قرار گرفته به‌طور معناداری عملکرد بهتری داشته است، در حالی که حرف w بیان‌گر عملکرد بدتر است. در وضعیتی که دو دسته‌بند اختلاف معناداری نداشته‌اند، حرفی درج نشده است.

[Downloaded from jhbmi.com on 2026-11-17]

جدول ۱: ویژگی‌های موجود در مجموعه داده آماده‌سازی شده

ردیف	نام ویژگی	سمبل	نوع داده	مقدار	توضیح	موجود/استخراجی
۱	جنسیت	Gender	Binary	۰ و ۱	زن، مرد	موجود
۲	سن	Age	Numeric	[۷۵، ۹۰]		استخراجی
۳	خاص بودن	Specialty	Binary	۰ و ۱	خاص، عادی	موجود
۴	محل صدور	Issue_Place	Binary	۰ و ۱	خارج از استان تهران و تهران	موجود
۵	نوع صندوق بیمه	Insur_Fund	Nominal	E، W، R و O	خویش‌فرما، کارکنان دولت، روستایی، سایر اقشار	موجود
۶	تعداد نسخه	Presc_Count	Numeric	≥۱		استخراجی
۷	تعداد پزشکان	Doc_Count	Numeric	≥۱		استخراجی
۸	میانگین تعداد نسخه هر پزشک	Presc_Doc_R	Numeric	≥۱		استخراجی
۹	درصد پزشکان زن مراجعه شده	Doc_Female_R	Numeric	[۰، ۱۰۰]		استخراجی
۱۰	درصد پزشکان مرد مراجعه شده	Doc_Male_R	Numeric	[۰، ۱۰۰]		استخراجی
۱۱	درصد پزشکان عمومی مراجعه شده	Doc_GP_R	Numeric	[۰، ۱۰۰]		استخراجی
۱۲	درصد پزشکان متخصص مراجعه شده	Doc_Prof_R	Numeric	[۰، ۱۰۰]		استخراجی
۱۳	درصد سایر پزشکان مراجعه شده	Doc_Oth_R	Numeric	[۰، ۱۰۰]		استخراجی
۱۴	درصد پزشکان با سطح تجربه ۱ مراجعه شده	Doc_E1_R	Numeric	[۰، ۱۰۰]	درصد پزشکان مراجعه شده با کد نظام پزشکی کوچک‌تر از ۴۰۰۰۰	استخراجی
۱۵	درصد پزشکان با سطح تجربه ۲ مراجعه شده	Doc_E2_R	Numeric	[۰، ۱۰۰]	درصد پزشکان مراجعه شده با کد نظام پزشکی بین ۴۰۰۰۰ و ۸۰۰۰۰	استخراجی
۱۶	درصد پزشکان با سطح تجربه ۳ مراجعه شده	Doc_E3_R	Numeric	[۰، ۱۰۰]	درصد پزشکان مراجعه شده با کد نظام پزشکی بین ۸۰۰۰۰ و ۱۲۰۰۰۰	استخراجی
۱۷	درصد پزشکان با سطح تجربه ۴ مراجعه شده	Doc_E4_R	Numeric	[۰، ۱۰۰]	درصد پزشکان مراجعه شده با کد نظام پزشکی بزرگ‌تر از ۱۲۰۰۰۰	استخراجی
۱۸	تعداد داروخانه‌های مراجعه شده	Store_Count	Numeric	≥۱		استخراجی
۱۹	میانگین نسخه‌های پیچیده شده در هر داروخانه	Presc_Store_R	Numeric	≥۱		استخراجی
۲۰	چندداروییی	PP	Binary	۰ و ۱	اگر میانگین تعداد داروی تجویزی در هر نسخه کمتر از ۳ باشد عدد صفر و در غیر این صورت یک خواهد بود.	استخراجی

جدول ۲: مقایسه دو به دو دسته‌بندهای مورد استفاده بر اساس شاخص حساسیت

	C4.5 (۰/۶۸(۰/۰۲)	BN (۰/۵۸(۰/۰۲)	SVM (۰/۷۰(۰/۰۲)	KNN (۰/۵۵(۰/۰۱)	MLP (۰/۶۳(۰/۲۰)
C4.5	-	b		b	
BN	w	-	W	b	
SVM		b	-	b	
KNN	w	w	W	-	
MLP					-

[Downloaded from jhbmi.com on 2026-11-17]

جدول ۳: مقایسه دو به دو دسته‌بندهای مورد استفاده بر اساس شاخص درستی

MLP	KNN	SVM	BN	C4.5	
۵۷/۴۹(۲/۴۷)	۵۲/۲۲(۰/۸۱)	۵۶/۰۶(۱/۰۳)	۵۸/۵۹(۱/۱۲)	۵۸/۱۳(۱/۱۵)	
	b	B		-	C4.5
	b	B	-		BN
	b	-	w	w	SVM
w	-	W	w	w	KNN
-	b				MLP

جدول ٤: مقایسه دو به دو دسته‌بندهای مورد استفاده بر اساس شاخص AUC

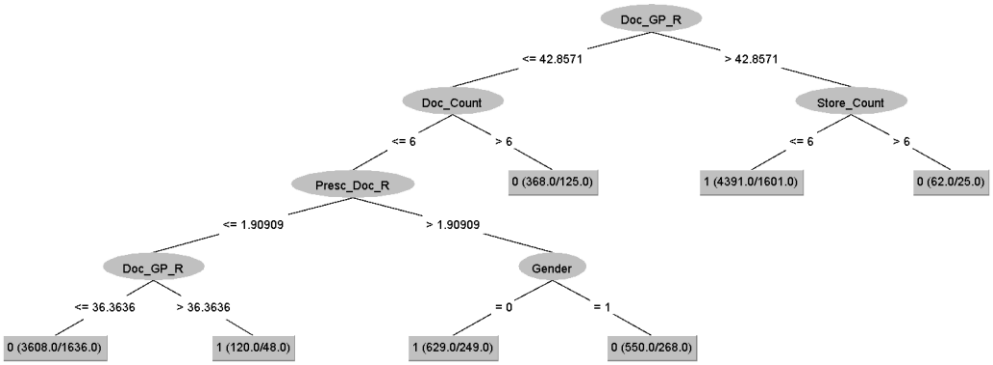
MLP	KNN	SVM	BN	C4.5	
۰/۶۱(۰/۰۱)	۰/۵۲(۰/۰۱)	۰/۵۵(۱/۰۱)	۰/۶۱(۰/۰۱)	۰/۵۸(۰/۰۱)	
w	b	B	w	-	C4.5
	b	B	-	b	BN
w	b	-	w	w	SVM
w	-	W	w	w	KNN
-	b	B		b	MLP

در ادامه و با توجه به این‌که از یک سو، C4.5 نسبت به سایر دسته‌بندهای مورد استفاده عملکرد بهتری را از خود به نمایش گذاشته است و از سوی دیگر، با ارائه یک ساختار درختی، توانایی ارائه مجموعه قوانینی را دارد که نشان‌گر ارتباط ویژگی‌های مختلف سالمندان بر پدیده چندداروییست، این الگوریتم با دقت بیشتری مورد بررسی قرار گرفته است. برای این منظور، مجموعه داده مورد مطالعه با تکنیک نمونه‌برداری طبقه‌بندی شده و با نسبت مساوی به دو زیر مجموعه تقسیم شده است، به‌طوری‌که در هر دو، نسبت نمونه‌ها با ویژگی چندداروییکسان است. سپس با در نظر گرفتن روش جستجوی توری و با در نظر گرفتن زیرمجموعه داده تنظیم (زیرمجموعه داده اول)، دو پارامتر مهم این دسته‌بند تنظیم شده است. مقدار بهینه پارامتر Confidence Factor در عدد ۰/۰۵۱ و مقدار بهینه پارامتر minNumObj در عدد ۱۳۵ به‌دست آمده است. این درحالی است که مقادیر پیش‌فرض این دو پارامتر به ترتیب ۰/۲۵ و ۲ می‌باشد که اختلاف فاحشی با مقدار بهینه تنظیمی دارد. مقایسه اجرای C4.5 در دو حالت با مقادیر پارامترهای پیش‌فرض و مقادیر پارامترهای تنظیم شده بر روی زیرمجموعه داده آزمون (زیرمجموعه داده دوم) در WEKA Experimenter انجام شده است، به‌طوری‌که هر یک از این الگوریتم‌ها هم‌زمان با اعتبارسنجی ضربدری ۱۰ بخشی، پنج بار بر روی زیرمجموعه آزمون اجرا شده‌اند و سپس آزمون زوجی t اصلاح شده با بازنمونه‌برداری در سطح معناداری ۹۵ درصد انجام شده است که حاکی از آن است که شاخص‌های عملکردی درستی و AUC برای C4.5 با تنظیم پارامتر به شکل معناداری بهبود یافته است؛ اما موجب کاهش در شاخص حساسیت شده است، ولی این کاهش در سطح ۹۵٪ معنادار نبوده است (جدول ۵). از سوی دیگر، تنظیم پارامتر موجب کاهش بسیار شدید اندازه درخت تصمیم خروجی شده است، به‌گونه‌ای که از یک درخت بسیار بزرگ با میانگین اندازه حدود ۹۵۸ گره و ۴۹۶ برگ (قانون) به درختی با میانگین اندازه ۱۵/۶۴ و میانگین تعداد برگ ۸/۳۴ رسیده است.

جدول ۵: شاخص‌های عملکردی C4.5 با مقادیر پارامترهای پیش‌فرض و تنظیم شده

	درستی	AUC	حساسیت	اندازه درخت	تعداد برگ (قوانین)
C4.5 (پیش فرض)	۵۶/۴۵(۱/۳۶)	۰/۵۶(۰/۰۲)	۰/۶۶(۰/۰۳)	۹۵۸/۰۸(۱۰۹/۲۵)	۴۹۶/۱۲(۵۶/۵۰)
C4.5 (تنظیم شده)	۵۹/۱۸(۱/۴۳)	۰/۶۰(۰/۰۲)	۰/۶۳(۰/۰۳)	۱۵/۶۴(۲/۱۲)	۸/۳۴(۱/۱۴)

[Downloaded from jhbmi.com on 2026-11-17]



شکل ۲: درخت تصمیم حاصل از C4.5 تنظیم پارامتر شده

در ادامه و به منظور ایجاد قوانین مورد نظر الگوریتم C4.5 با در نظر مقادیر تنظیمی پارامترها در کاربرگ Explorer نرم‌افزار WEKA بر روی زیرمجموعه آزمون اجرا شده است. درخت تصمیم خروجی در شکل ۲ نشان داده شده است. همان‌گونه که مشاهده می‌شود این درخت دارای ۱۳ گره و ۷ برگ (قانون) می‌باشد که در این میان تنها سه برگ در بردارنده چند دارویی است (مستطیل‌هایی که عدد ۱ در آن‌ها درج شده است). به‌بیان‌دیگر، نحوه شناسایی چندارویی تنها با سه قانون قابل بیان است. هر چند درخت تصمیم تولید شده به دلیل کوچک بودن از گویایی بالایی برخوردار است، اما برای درک بهتر برای مدیران و تصمیم‌گیرندگان، این سه قانون در قالب عبارات کلامی به صورت زیر ارائه داد.

۱- بیمارانی که درصد بیشتری از نسخه‌های آن‌ها توسط پزشکان عمومی تجویز شده است (بیش از حدود ۴۲/۸۶٪) و داروهایشان را از تعداد محدودتری از داروخانه‌ها گرفته‌اند (کمتر از ۷ داروخانه) به‌طور میانگین بیش از سه دارو در هر نسخه آن‌ها تجویز شده است.

۲- بانوانی که درصد کمتری از نسخه‌های آن‌ها توسط پزشکان عمومی تجویز شده است (کمتر از حدود ۴۲/۸۶٪)، اما به‌طور میانگین مراجعات متعددی به پزشکان داشته‌اند (بیشتر از ۱/۹ مراجعه) و در مجموع کمتر از ۷ نسخه داشته‌اند، تمایل به چندارویی داشته‌اند.

۳- بیمارانی که درصد نسبتاً کمتری از نسخه‌های آن‌ها توسط پزشکان عمومی تجویز شده است (بین ۳۶/۳۶٪ و ۴۲/۸۶٪)، و حتی به‌طور میانگین مراجعات متعددی به پزشکان نداشته‌اند (کمتر از ۱/۹ مراجعه) و در مجموع کمتر از ۷ نسخه داشته‌اند، نیز دچار چندارویی شده‌اند.

بحث و نتیجه‌گیری

بررسی‌های انجام شده در این پژوهش نشان‌گر آن است که داده‌های نسخه‌های تجویزی به تنهایی قابلیت داده‌کاوی را ندارند و لازم است تا با سایر منابع اطلاعات از جمله اطلاعات پزشکان و نیز اطلاعات تکمیلی بیماران یکپارچه شوند. همچنین لازم است تا کیفیت داده‌های جمع‌آوری مورد بررسی دقیق‌تر و گسترده‌تری قرار گیرد. فراتر از این لازم است تا ویژگی‌های جدیدی از داده‌ها استخراج شود. در این پژوهش از ۲۰ ویژگی مجموعه داده آماده‌سازی شده برای داده‌کاوی تنها ۴ ویژگی در مجموعه داده اولیه وجود داشته است.

از سوی دیگر، عملکرد دسته‌بندهای مختلف بر روی مجموعه داده آماده‌سازی شده، یکسان نبوده است. مقایسه زوجی بر مبنای شاخص حساسیت که توانایی الگوریتم‌ها را در شناسایی سالمندانی که دچار چندارویی شده‌اند را مشخص می‌کند نشان‌گر شرایط بهتر SVM و C4.5 در سطح معناداری ۹۵٪ است؛ اما اگر مبنای مقایسه شاخص درستی باشد، آنگاه BN و C4.5 به‌طور معناداری عملکرد بهتری داشته‌اند. این در حالی است که بر مبنای AUC، دو دسته‌بند BN و MLP با عملکرد مشابه در صدر قرار دارند و بعد از آن‌ها C4.5 قرار دارد. الگوریتم KNN در هر سه شاخص در سطح اطمینان ۹۵٪ نسبت به سایر دسته‌بندها عملکرد ضعیف‌تری داشته است. در مجموع می‌توان عملکرد C4.5 را در مقایسه با سایر دسته‌بندها مطلوب ارزیابی کرد. همچنین توجه به این نکته حائز اهمیت است که همگی دسته‌بندها به جزء C4.5 به‌صورت جعبه سیاه عمل می‌کنند. به بیان دیگر، تنها C4.5 قابلیت نمایش نحوه تشخیص برچسب ویژگی هدف را داراست. این کار از طریق ایجاد درخت تصمیم و نیز ارائه مجموعه قوانینی در قالب اگر-آن‌گاه (If-Then) انجام می‌شود. مقایسه C4.5 با مقادیر پارامترهای پیش‌فرض و مقادیر پارامترهای تنظیم شده حکایت از بهبود معنادار شاخص‌های عملکردی

[Downloaded from jhbmi.com on 2026-11-17]

الگوریتم C4.5 در اثر تنظیم پارامتر دارد. در واقع تنظیم پارامتر به‌وسیله الگوریتم جستجوی توری ضمن ارتقاء هم‌زمان دو شاخص درستی و AUC موجب تولید درخت تصمیم به مراتب کوچک‌تری شده است که قوانین حاصل از آن بسیار ساده‌تر و قابل فهم‌تر است.

یافته‌های این پژوهش نشان‌گر آن است که استفاده از رویکرد داده‌کاوی و به‌خصوص به‌کارگیری C4.5 توانایی شناسایی ویژگی‌های بیماران سالمند را بر بروز چندداروییی داشته است. در این میان، بیمارانی که درصد بیشتری از مراجعاتشان به پزشکان عمومی بوده است (بیش از حدود ۴۳ درصد) و با تعداد محدودتری از داروخانه‌ها در ارتباط بوده‌اند (کمتر از ۷ داروخانه) به‌طور میانگین بیش از سه دارو در هر نسخه آن‌ها تجویز شده است. این قانون به تنهایی ۲،۷۹۰ نفر از کل ۵،۲۹۶ نفری که دچار چندداروییی شده‌اند را شناسایی کرده است که نشان‌گر شناسایی حدود ۵۳٪ از موارد چندداروییی با اطمینان بیش از ۶۳٪ است. در مجموع درصد نسخه‌هایی که توسط پزشکان عمومی تجویز شده است، به‌عنوان یک ویژگی کلیدی مطرح است. علاوه بر این تعداد داروخانه‌های مرتبط، تعداد نسخه‌های تجویزی، تعداد مراجعات مکرر به پزشکان و جنسیت بیمار بر تجویز بیش‌ازحد دارو در هر نسخه و بروز چندداروییی مؤثر بوده است.

به منظور بهبود قوانین استخراجی برای تبیین عوامل مؤثر بر چند دارویی و ارتقاء کیفیت نتایج به‌دست آمده، لازم است تا کیفیت داده‌های اولیه پژوهش ارتقاء یابد. این کار می‌تواند با ایجاد تغییر در نحوه دریافت اطلاعات از داروخانه‌ها و ذخیره‌سازی داده‌ها در اداره‌های بیمه سلامت صورت پذیرد و یا اطلاعات بیشتری در خصوص ویژگی‌های دیگری از بیماران (مانند سطح تحصیلات، قومیت، وضعیت اقتصادی-اجتماعی و ...) جمع‌آوری شود. علاوه بر این می‌توان از روش‌های دیگری برای تنظیم هم‌زمان تعداد بیشتری از پارامترها استفاده کرد و نتایج را با یافته‌های این پژوهش مقایسه کرد. همچنین نویسندگان این پژوهش، به اطلاعات مصرف خودسرانه داروهای بدون نسخه (Over the Counter) OTC و یا داروهای مکمل از سوی بیمار دسترسی نداشته است. علاوه بر این، در صورتی که بیماران در طول سال ۱۳۹۴ در بیمارستانی بستری شده باشند و برای آن‌ها داروهایی تجویز شده باشد، از آنجا که اطلاعات آن در اختیار نبوده است، این داده‌ها در این پژوهش لحاظ نشده است.

نتایج حاصل از این پژوهش می‌تواند با شناسایی برخی ویژگی‌های بیماران سالمند که در بروز چندداروییی مؤثر است، تصویر شفاف‌تر و دقیق‌تری در اختیار سیاست‌گذاران و مدیران حوزه سلامت و به‌خصوص صنعت بیمه قرار دهد تا بتوانند اقدامات متناسبی را برای منطقی‌سازی تجویز دارو انجام دهند و از این طریق بتوانند تجویز و به دنبال آن مصرف دارو را برای سالمندان کاهش دهند. همچنین روش داده‌کاوی به‌کار رفته در این پژوهش می‌تواند به‌عنوان بخشی از یک رویکرد مکانیزه در فرآیند واکاوی منطقی بودن نسخه‌های تجویزی برای بیماران به کار گرفته شود.

تشکر و قدردانی

نویسندگان بر خود لازم می‌دانند تا برای تأمین داده‌های اولیه پژوهش از سازمان بیمه سلامت ایران، به‌خصوص مرکز ملی تحقیقات بیمه سلامت و همچنین اداره کل بیمه سلامت استان تهران کمال قدردانی را داشته باشند. علاوه بر این، سپاس ویژه خود را به جناب دکتر شجاعی، خانم شاهعلی و خانم مهندس عرب و همچنین مهندس اقلیدی که انجام این پژوهش بدون همکاری و همیاری ایشان میسر نبود، اعلام می‌دارند.

تعارض منافع

در انجام این پژوهش، هیچ‌گونه تعارض منافعی وجود ندارد. این مقاله حاصل رساله دکتری در دانشکده مهندسی صنایع دانشگاه صنعتی شریف است.

References

1. Finkers F, Maring JG, Boersma F, Taxis K. A study of medication reviews to identify drug-related problems of polypharmacy patients in the Dutch nursing home setting. J Clin Pharm Ther 2007;32(5):469-76. doi: 10.1111/j.1365-2710.2007.00849.x

2. Fulton MM, Riley Allen E. Polypharmacy in the elderly: a literature review. J Am Acad Nurse Pract 2005;17(4):123-32. doi: 10.1111/j.1041-2972.2005.0020.x

3. Hovstadius B, Petersson G. Factors leading to excessive polypharmacy. Clin Geriatr Med 2012;28(2):159-72. doi: 10.1016/j.cger.2012.01.001

[Downloaded from jhbmi.com on 2026-11-17]

4. Huang AR, Mallet L, Rochefort CM, Eguale T, Buckeridge DL, Tamblyn R. Medication-related falls in the elderly. Drugs Aging 2012;29(5):359-76. doi: 10.2165/11599460-000000000-00000
5. Galazzi A, Lusignani M, Chiarelli MT, Mannucci PM, Franchi C, Tettamanti M, Reeve E, Nobili A. Attitudes towards polypharmacy and medication withdrawal among older inpatients in Italy. Int J Clin Pharm 2016;38(2):454-61. doi: 10.1007/s11096-016-0279-4
6. Balkhi B, AlQahtani N, Alwhaibi M, Alshammari TM, Alhawassi TM, Mahmoud MA, et al. Prevalence and factors associated with polypharmacy use among adult patients in Saudi Arabia. J Patient Saf 2017. doi: 10.1097/PTS.0000000000000439
7. Morin L, Johnell K, Laroche ML, Fastbom J, Wastesson JW. The epidemiology of polypharmacy in older adults: register-based prospective cohort study. Clin Epidemiol 2018; 10: 289–98. doi: 10.2147/CLEP.S153458
8. Rieckert A, Trampisch US, Klaaßen-Mielke R, Drewelow E, Esmail A, Johansson T, et al. Polypharmacy in older patients with chronic diseases: a cross-sectional analysis of factors associated with excessive polypharmacy. BMC Family Practice 2018;19(1):113. doi: 10.1186/s12875-018-0795-5
9. Delshad Noghabi A, Baloochi Beydokhti T, Shamshiri M, Shareinia H, Radmanesh R. Polypharmacy and its related factors among elderlies. Iran Journal of Nursing 2013; 26(83):1-9. [In Persain]
10. Dianati M, Shojaegharebag G, Mesdaghinia A, Taghadosi M, Shenasa F, Taiebi A, et al. Polypharmacy and its related factors among the elderly population in Kashan, Iran during 2011-2012. Feyz 2015;18(6): 578-84. [In Persian]
11. Torabi B, Mohammadi M, Aboozade Gatabi K, Ghanbari Moghaddam A. study of prescribing unsuitable drugs and polypharmacy in the elderly admitted to the Vasei Hospital of Sabzevar, 2017. Journal of Gerontology 2018;3(1):48-55. [In Persian] doi: 10.29252/joge.3.1.48
12. Held F, Le Couteur DG, Blyth FM, Hirani V, Naganathan V, Waite LM, et al. Polypharmacy in older adults: Association Rule and Frequent-Set Analysis to evaluate concomitant medication use. Pharmacol Res 2017;116:39-44. doi: 10.1016/j.phrs.2016.12.018
13. Moradi M, Modarres M, Sepehri MM. Analysis of the Average Number of Drugs Prescribed Per Encounter by Internists: A Data Mining Approach. 10th Iran Data Mining Conference; 2018 Jan 16; Tehran: 2018. [In Persian]
14. Moradi M, Modarres M, Sepehri MM. Analysis of the Polypharmacy in psychiatrist Perceptions: data mining of imbalanced dataset. 15th International Industrial Engineering Conference, 2019 Jan 23-24; Yazd: Yazd University; 2019. [In Persian]
15. Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3rd ed. USA: Morgan Kaufmann; 2011.
16. Larose DT, Larose CD. Discovering knowledge in data: an introduction to data mining. New York: John Wiley & Sons; 2014.
17. Fu W, Nair V, Menzies T. Why is differential evolution better than grid search for tuning defect predictors?. arXiv preprint arXiv:1609.02613. 2016; 1-13.
18. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. The Journal of Machine Learning Research 2012;13(1):281-305.
19. Barat I, Andreasen F, Damsgaard EM. The consumption of drugs by 75-year-old individuals living in their own homes. Eur J Clin Pharmacol 2000;56(6-7):501-9. doi: 10.1007/s002280000157
20. Alic A, Pranjic N, Ramic E. Polypharmacy and decreased cognitive abilities in elderly patients. Med Arh 2011;65(2):102-5.
21. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatrics 2017;17(1):230. doi: 10.1186/s12877-017-0621-2

[Downloaded from jhbmi.com on 2026-11-17]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2020; 7(2): 150-160

Comparison of Classifier Algorithms in the Identification of Polypharmacy and Factors Affecting it in the Elderly Patients

Moradi Morteza[1]*, Modarres Mohammad[2], Sepehri Mohammad Mehdi[3]

• **Received:** 09 Aug 2019 • **Accepted:** 16 Sep 2019

Introduction: Prescribing and consuming drugs more than necessary which is known as polypharmacy, is both waste of resources and harm to patients. Polypharmacy is especially important for elderly patients; therefore, the factors affecting it must be identified and analyzed properly.

Method: In this retrospective study, first, several classifier algorithms, i.e., C4.5, SVM, KNN, MLP, and BN for polypharmacy identification were compared in terms of performance using WEKA software. In this process, 16 new features were extracted alongside the four existing features from data on 81,677 prescriptions of 19,428 outpatients aged 70 to 95 years whose prescriptions were dispensed in pharmacies contracted by the Iran Health Insurance Organization-Tehran province. The performance comparison was done using corrected t-test with resampling. In order to identify the effect of elderly patients' characteristics on polypharmacy, two important parameters of the C4.5 were optimized by grid search using 50% of the dataset and then run on the rest of the dataset. The resulted rules were then presented in the form of a decision tree and verbal expressions.

Results: Paired comparison of the classifiers indicated better performance of C4.5 and BN compared to the others. C4.5 had the ability to identify the factors that affect polypharmacy. In addition, parameter tuning improved the accuracy and AUC of applied algorithms. It also reduced the size of the resulted decision trees as well as the number of generated rules significantly.

Conclusion: The data mining approach and C4.5 can identify and explain the characteristics of the elderly effective on the polypharmacy. The higher percentage of visits to general practitioners and contacts with a limited number of pharmacies are the most important characteristics.

Keywords: Data Mining, Classification, Polypharmacy, Rational Prescription and Use of Drugs, Elderly

• **Citation:** Moradi M, Modarres M, Sepehri MM. Comparison of Classifier Algorithms in the Identification of Polypharmacy and Factors Affecting it in the Elderly Patients. Journal of Health and Biomedical Informatics 2020; 7(2): 150-60. [In Persian]

1. Ph.D. Student in Industrial Engineering, School of Industrial Engineering, Sharif University of Technology, Tehran, Iran
2. Ph.D. in Systems Engineering and Operations Research, Professor, School of Industrial Engineering, Sharif University of Technology, Tehran, Iran
3. Ph.D. in Operations Research, Professor, School of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran

*__Corresponding Author__: Morteza Moradi
Address: School of Industrial Engineering, Sharif University of Technology, Azadi Ave., Tehran, Iran
• **Tel:** +98-21-66165719 • **Email:** m.moradi@ie.sharif.edu

[Downloaded from jhbmi.com on 2026-11-17]