Research Article

The Iranian Journal of Health Insurance | Volume 9, Issue 1, Spring 2026 | Pages: 13-22

journal.ihio.gov.ir

Exploration of Co Prescribing Patterns of Medications in Alzheimer's Patients Using Association Rule Mining on Real Health Insurance Data in Iran (2020–2022)

Maedeh Gholamazad[1] PhD, Alireza Atashi[1] PhD, Fatemeh Hajialiasgari[2] MD, Shima Mohammadi[3,4*] PhD, Reza Mehrizi[5] MD, Leila Ghamkhar[3] PhD, Hossein Effatpanah[6] PhD

[1] Medical Informatics Department, Breast Cancer Research Center, National Cancer Institute-Iran, ACECR, Tehran, Iran
[2] Geriatrics Medicine Department, School of Medicine, Tehran University of Medical Sciences, Iran
[3] National Center for Health Insurance Research, Tehran, Iran
[4] Neuroscience Research Center, Shahroud University of Medical Sciences, Shahroud, Iran
[5] School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[6] Department of School Health, Asadabad faculty of Medical Sciences, Asadabad, iran

***Correspondence to:** Shima Mohammadi, Email: shimamohamadi1365@gmail.com

Received: March 25,2026 **Revised:** May 19, 2026 **Accepted:** May 31, 2026 **Online Published:** June 8, 2026

Abstract

Introduction: Alzheimer's disease, as the most common cause of dementia in older adults, imposes a substantial burden on the healthcare system. With the growing elderly population in Iran, understanding real world prescribing patterns is essential for optimizing treatment and reducing polypharmacy. Association rule mining on insurance registry data provides an opportunity to identify co occurrence patterns of medications.

Methods: A descriptive–analytical, non experimental study was conducted on registry data from the Iran Health Insurance Organization (2020–2022). From 532,369 prescription records of Alzheimer's patients, after cleaning, standardization, and removal of rare medications, a binary transactional matrix (presence/absence of drugs) was constructed. The Apriori and FP Growth algorithms were applied to extract association rules using Support, Confidence, and Lift.

Results: Donepezil and Memantine dominated with 81% of prescriptions; Rivastigmine and Galantamine had minimal shares, and non specific medications were limited. Both algorithms showed similar patterns (FP Growth more efficient), yet no meaningful rules met the Support, Confidence, and Lift thresholds; convergence of results indicated high dispersion in real world data.

Conclusion: Prescribing patterns in Alzheimer's disease were consistent with clinical guidelines but lacked stable co occurrences for classical association rules. These findings highlight the complexity of clinical data and the need for Sequential Mining and machine learning, providing a foundation for optimizing prescribing and policymaking in Iran.

Keywords: Alzheimer's Disease, Drug Prescribing, Data Mining, Association Rules, Apriori, FP Growth, Health Insurance

Highlights:

1. This study analyzed nationwide real-world prescribing patterns in Alzheimer's disease using Iran Health Insurance registry data and the Apriori and FP-Growth association rule mining algorithms.
2. Despite the large dataset, no stable association rules were identified, highlighting the complexity of prescribing behaviors and the need for advanced approaches such as sequential mining and machine learning.

Citation:
Gholamazad M, Atashi A, Hajialiasgari F, Mohammadi Sh, Mehrizi R, Ghamkhar L, et al. Exploration of Co Prescribing Patterns of Medications in Alzheimer's Patients Using Association Rule Mining on Real Health Insurance Data in Iran (2020–2022). Iran J Health Insur. 2026;9(1):13-22.

Copyright © 2026 The Author(s); Published by National Center for Health Insurance Research. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[DOI: 10.66224/ihio.9.1.6]

مقاله پژوهشی

فصلنامه علمی- پژوهشی بیمه سلامت ایران | دوره ۹، شماره ۱، بهار ۱۴۰۵ | صفحات ۱۳ تا ۲۲

کاوش الگوهای هم‌وقوعی داروهای تجویزی در بیماران آلزایمر با استفاده از قوانین انجمنی بر روی داده‌های واقعی بیمه سلامت ایران (۱۳۹۹-۱۴۰۱)

مائده غلام‌آزاد[۱] PhD، علیرضا آتشی[۱] PhD، فاطمه حاجی‌علی‌عسگری[۲] MD، شیما محمدی[۳،۴*] PhD، رضا مهریزی[۵] MD، لیلا غمخوار[۳] PhD، حسین عفت‌پناه[۶] PhD

[۱] گروه پژوهشی هوش مصنوعی، انفورماتیک و مدل‌سازی پزشکی، مرکز تحقیقات سرطان پستان، پژوهشگاه ملی سرطان، جهاد دانشگاهی، تهران، ایران
[۲] گروه طب سالمندی، دانشکده پزشکی، دانشگاه علوم پزشکی تهران، ایران
[۳] مرکز ملی تحقیقات بیمه سلامت، تهران، ایران
[۴] مرکز تحقیقات علوم اعصاب، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران
[۵] دانشکده پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
[۶] گروه بهداشت مدارس،دانشکده علوم پزشکی اسدآباد،اسدآباد،ایران.

* **نویسنده مسئول:** شیما محمدی، پست الکترونیک: shimamohamadi1365@gmail.com

دریافت: ۱۴۰۵/۰۱/۰۵ تصحیح: ۱۴۰۵/۰۲/۲۹ پذیرش: ۱۴۰۵/۰۳/۱۰ انتشار آنلاین: ۱۴۰۵/۰۳/۱۸

چکیده

مقدمه: بیماری آلزایمر به‌عنوان شایع‌ترین علت زوال عقل در سالمندان، بار قابل‌توجهی بر نظام سلامت تحمیل می‌کند. با روند رو به افزایش جمعیت سالمند در ایران، شناخت الگوهای تجویز دارو در شرایط واقعی بالینی برای بهینه‌سازی درمان و کاهش پولی‌فارماسی ضروری است. کاوش قوانین انجمنی بر داده‌های ثبتی بیمه‌ای فرصتی برای شناسایی الگوهای هم‌وقوعی داروها فراهم می‌آورد.

روش بررسی: مطالعه توصیفی-تحلیلی و غیرتجربی بر داده‌های ثبتی سازمان بیمه سلامت ایران (۱۳۹۹-۱۴۰۱) انجام شد. از ۵۳۲۳۶۹ رکورد نسخه‌ای بیماران آلزایمری پس از پاک‌سازی، استانداردسازی و حذف داروهای نادر، ماتریس تراکنشی دودویی (حضور/عدم حضور دارو) ساخته شد. الگوریتم‌های Apriori و FP-Growth برای استخراج قوانین انجمنی با معیارهای Support، Confidence و Lift اجرا گردیدند.

یافته‌ها: Donepezil و Memantine با تمرکز ۸۱ درصد تجویزها غالب بودند؛ Rivastigmine و Galantamine سهم ناچیزی داشتند و داروهای غیراختصاصی محدود بودند. هر دو الگوریتم الگوهای مشابهی نشان دادند (FP-Growth کارایی بالاتر)، اما هیچ قانون معناداری با آستانه‌های Support، Confidence و Lift شناسایی نشد؛ همگرایی نتایج بیانگر پراکندگی بالای داده‌های واقعی است.

نتیجه‌گیری: الگوهای تجویزی آلزایمر با راهنماهای بالینی همخوان اما فاقد هم‌وقوعی پایدار برای قوانین انجمنی کلاسیک است. این یافته‌ها پیچیدگی داده‌های بالینی را تأیید و روش‌های Sequential Mining و یادگیری ماشین را ضروری می‌سازد؛ مبنایی برای بهینه‌سازی تجویز و سیاست‌گذاری در ایران فراهم می‌کند.

واژگان کلیدی: بیماری آلزایمر، تجویز دارو، داده‌کاوی، قوانین انجمنی، FP-Growth، Apriori، نسخه‌های بیمه سلامت

نکات ویژه

۱- این مطالعه با استفاده از داده‌های واقعی بیمه سلامت ایران و الگوریتم‌های Apriori و FP-Growth، الگوهای تجویز دارویی بیماران مبتلا به آلزایمر را در سطح ملی بررسی کرد.

۲-عدم وجود قوانین انجمنی پایدار علی‌رغم حجم بالای داده‌ها، پیچیدگی الگوهای تجویز در بیماران آلزایمری را نشان داد و ضرورت استفاده از روش‌های پیشرفته‌تر مانند یادگیری ماشین را برجسته ساخت.

مقدمه

بیماری آلزایمر (AD) شایع‌ترین علت زوال عقل در سالمندان است و به‌عنوان یک اختلال عصب‌تخریبی پیش‌رونده، با کاهش تدریجی توانایی‌های شناختی، اختلال حافظه کوتاه‌مدت، تغییرات رفتاری و افت قابل توجه کیفیت زندگی مشخص می‌شود که در نهایت می‌تواند به وابستگی کامل فرد به مراقب و مرگ زودرس منجر گردد [۱]. علت اصلی این بیماری هنوز به‌طور کامل

[DOI: 10.66224/ihio.9.1.6]

شناخته نشده است، اما شواهد حاکی از نقش مهم سن بالا، زمینه ژنتیکی و عوامل خطر غیرژنتیکی (مانند بیماری‌های عروقی، سبک زندگی غیرفعال و سطح تحصیلات پایین) در بروز و پیشرفت آن است؛ به‌طوری که خطر ابتلا پس از ۶۵ سالگی تقریباً هر پنج سال دو برابر می‌شود و بیماری معمولاً در سه مرحله بالینی خفیف، متوسط و شدید سیر می‌کند [۲]. برآوردها نشان می‌دهد که حدود ۳۵/۶ میلیون نفر در سال ۲۰۱۰ با دمانس زندگی می‌کرده‌اند [۳] و انتظار می‌رود این تعداد تا سال ۲۰۵۰ به حدود ۱۱۵/۴ میلیون نفر برسد، به‌ویژه در کشورهای کم‌درآمد و متوسط که با پدیده سالمندی جمعیت مواجه‌اند. پیش‌بینی‌ها نشان می‌دهد جمعیت سالمندان بالای ۶۰ سال در ایران از حدود ۹,۱میلیون نفر در سال ۲۰۲۰ به حدود ۱۳,۷ میلیون نفر در سال ۲۰۲۹ افزایش خواهد یافت. همزمان، تعداد بیماران مبتلا به آلزایمر نیز روندی صعودی دارد، به‌طوری‌که انتظار می‌رود از حدود ۴۶۴٬۴۰۰ نفر در سال ۲۰۲۰ به حدود ۷۲۹٬۹۰۰ نفر در سال ۲۰۲۹ برسد [۴].

در سال‌های اخیر، درمان دارویی دمانس و به‌ویژه آلزایمر در مرکز توجه پژوهش‌های علوم اعصاب و پزشکی قرار گرفته است، هرچند درمان‌های موجود عمدتاً ماهیت علامتی داشته و بر بهبود نسبی عملکرد شناختی و کنترل علائم رفتاری متمرکز هستند. دو گروه دارویی اصلی در مدیریت آلزایمر شامل مهارکننده‌های کولین‌استراز (ChEIs) و آنتاگونیست‌های گیرنده N-methyl-D-aspartate هستند که به‌عنوان درمان‌های استاندارد برای مراحل خفیف تا متوسط (ChEIs) و متوسط تا شدید ممانتین توصیه می‌شوند. داروهایی مانند دونپزیل، ریواستیگمین و گالانتأمین با مهار آنزیم استیل‌کولین‌استراز، از تجزیه استیل‌کولین در سیناپس‌های مغزی جلوگیری کرده و با افزایش غلظت آن در شکاف سیناپسی، انتقال عصبی در مسیرهای کولینرژیک قشر مغز و هیپوکامپ را تقویت می‌کنند؛ کاهش عملکرد کولینرژیک یکی از یافته‌های ثابت در پاتوفیزیولوژی آلزایمر است و تقویت این سیستم می‌تواند به بهبود نسبی حافظه، توجه و سایر کارکردهای شناختی منجر شود [۵]. همزمان، ممانتین به‌عنوان یک آنتاگونیست غیررقابتی و با میل پایین به گیرنده NMDA، با تعدیل فعالیت گلوتامات و جلوگیری از تحریک‌پذیری سمی، نقش مهمی در مراحل متوسط تا شدید آلزایمر ایفا می‌کند و اغلب در ترکیب با مهارکننده‌های کولین‌استراز تجویز می‌شود تا اثرات شناختی و عملکردی بهتری حاصل گردد [۶].

با وجود تدوین راهنماهای بالینی و تعریف داروهای استاندارد برای درمان آلزایمر، شواهد نشان می‌دهد که الگوهای تجویز در عمل بالینی می‌تواند تحت تأثیر عوامل متعددی مانند سن، جنس، وضعیت اجتماعی–اقتصادی، هم‌ابتلایی‌ها، ترجیحات پزشک و دسترسی به خدمات سلامت قرار گیرد. در بسیاری از کشورها، از داده‌های ثبتی بیمه‌ای برای توصیف الگوهای تجویز داروهای ضددمانس و بررسی نابرابری در دسترسی بیماران به این داروها استفاده شده است؛ نتایج این مطالعات نشان می‌دهد که تفاوت‌های معناداری در نوع داروی انتخابی، شدت و مرحله بیماری هدف بین گروه‌های مختلف جمعیتی وجود دارد [۷]. با این حال، در ایران، علی‌رغم حجم قابل توجه داده‌های نسخه‌ای در سامانه‌های بیمه سلامت، کاوش نظام‌مند الگوهای تجویز دارو در بیماران مبتلا به آلزایمر با رویکردهای پیشرفته داده‌کاوی کمتر انجام شده و عمده مطالعات موجود به گزارش‌های توصیفی محدود شده‌اند. این خلأ پژوهشی، به‌ویژه در زمینه تحلیل الگوهای هم‌وقوعی داروها و شناسایی ترکیبات تجویزی رایج یا غیرمنطقی، می‌تواند منجر به تداوم تجویزهای غیربهینه و افزایش هزینه‌های مستقیم و غیرمستقیم بیماری شود.

یکی از رویکردهای قدرتمند برای تحلیل الگوهای پیچیده تجویز در داده‌های بزرگ‌مقیاس نسخه‌ای، استفاده از روش‌های کاوش قوانین انجمنی مانند الگوریتم‌های Apriori و FP-Growth است. در این روش‌ها، داده‌ها به صورت تراکنشی مدل‌سازی می‌شوند؛ به این معنا که هر

[DOI: 10.66224/ihio.9.1.6]

بیمار (یا هر دوره درمانی) به‌عنوان یک «تراکنش» و هر داروی تجویزی به‌عنوان یک «آیتم» در نظر گرفته می‌شود و سپس روابطی از نوع $X \rightarrow Y$ که در آن X (اقلام مقدم) و Y (اقلام متعاقب) هستند، بر اساس معیارهایی مانند پشتیبانی، اطمینان و ضریب بالابری استخراج می‌گردند. این معیارها به‌ترتیب میزان فراوانی هم‌وقوعی، قدرت پیش‌بینی و میزان انحراف از استقلال تصادفی را نشان می‌دهند و می‌توانند برای شناسایی ترکیبات دارویی که به‌طور پایدار و معنادار در کنار هم تجویز می‌شوند، مورد استفاده قرار گیرند [۸]. علی‌رغم کاربرد گسترده این روش‌ها در داده‌کاوی پزشکی، گزارش‌ها نشان داده‌اند که در داده‌های با ابعاد بالا و ناهمگن، بخش قابل توجهی از قوانین استخراج‌شده دارای پشتیبانی بسیار پایین بوده و فاقد ارزش بالینی هستند؛ بنابراین نیاز به تنظیم دقیق آستانه‌ها و تفسیر محتاطانه نتایج وجود دارد [۹].

بر این اساس، هدف این مطالعه کاوش الگوهای هم‌وقوعی داروهای تجویزی در بیماران مبتلا به آلزایمر، با استفاده از روش‌های کاوش قوانین انجمنی (Apriori و FP-Growth) بر روی داده‌های واقعی نسخه‌ای بیمه سلامت ایران در سال‌های ۱۳۹۹ تا ۱۴۰۱ است؛ به‌گونه‌ای که ضمن توصیف فراوانی و ترکیب داروهای اختصاصی و غیراختصاصی، الگوهای تجویزی پایدار شناسایی شده و زمینه برای بهینه‌سازی تجویزها و سیاست‌گذاری مبتنی بر شواهد فراهم شود.

روش بررسی

در این مطالعه از نوع توصیفی–تحلیلی و مدلسازی، از داده‌های ثبتی ثانویه مربوط به نسخه‌های دارویی بیماران مبتلا به آلزایمر استفاده شد. جامعه مورد مطالعه، کلیه بیماران دارای تشخیص آلزایمر ثبت‌شده در سامانه سازمان بیمه سلامت ایران بود که نسخه‌های دارویی آنها در بازه زمانی ابتدای سال ۱۳۹۹ تا پایان سال ۱۴۰۱ در سطح کشور ثبت شده است. داده‌ها به‌صورت یکپارچه از بانک اطلاعاتی الکترونیک نسخ بیمه سلامت استخراج شده و مطالعه به‌صورت غیرتجربی انجام گردید. دامنه پژوهش شامل داروهای تجویزی مرتبط و غیرمرتبط با آلزایمر و هدف اصلی، شناسایی الگوهای تجویز و هم‌وقوعی داروها در شرایط واقعی بالینی بود.

سوژه‌ها، ملاک‌های ورودو خروج و حجم داده

واحد تحلیل در این پژوهش، «پرونده نسخه‌ای بیمار مبتلا به آلزایمر» بود. ملاک ورود شامل: ۱) وجود کد تشخیصی آلزایمر (مطابق کدهای ثبت‌شده در سامانه بیمه سلامت) در پرونده بیمار و ۲) داشتن حداقل یک نسخه دارویی ثبت‌شده در بازه ۱۳۹۹ تا ۱۴۰۱. پرونده‌هایی که تشخیص آلزایمر در آنها نامشخص یا ناقص ثبت شده بود، یا فاقد اطلاعات دارویی معتبر بودند، از مطالعه حذف شدند. با توجه به ماهیت ثبتی و در دسترس بودن کل جمعیت هدف، نمونه‌گیری انجام نشد و تمامی پرونده‌های واجد شرایط در تحلیل وارد شدند؛ در مجموع، حدود ۵۳۲٬۳۶۹ رکورد نسخه‌ای مرتبط با بیماران آلزایمری مورد بررسی قرار گرفت. این حجم بالای داده امکان مدلسازی الگوهای تجویز در سطح جمعیتی را فراهم کرد.

نحوه جمع‌آوری و آماده‌سازی داده‌ها

داده‌ها به‌صورت فایل‌های الکترونیکی استاندارد از سامانه سازمان بیمه سلامت دریافت و پس از پاک‌سازی اولیه، برای تحلیل آماده‌سازی شدند. هر رکورد شامل شناسه ناشناس بیمار، تاریخ نسخه، فهرست داروهای تجویز شده بود. برای حفظ محرمانگی، تمامی شناسه‌های فردی پیش از تحلیل به‌صورت کدگذاری‌شده و غیرقابل رجعت در اختیار تیم پژوهش قرار گرفت و هیچ‌گونه متغیر هویتی (نام، کد ملی، شماره تماس) در داده نهایی وجود نداشت. سپس داده‌ها براساس شناسه بیمار و دوره درمانی تجمیع شدند به‌گونه‌ای که هر بیمار (یا هر دوره درمانی مشخص) به‌عنوان یک «تراکنش» و مجموعه داروهای

[DOI: 10.66224/ihio.9.1.6]

تجویزی در آن دوره به‌عنوان «آیتم‌ها» در نظر گرفته شد. رکوردهای ناقص، نسخه‌های با کد دارویی نامعتبر و موارد تکراری در مرحله پیش‌پردازش حذف شدند.

ابزارها و مراحل پردازش داده

برای مدیریت و تحلیل داده‌ها از زبان برنامه‌نویسی Python (نسخه ۳,۱۴,۴) و کتابخانه‌های Pandas و Numpy برای پردازش داده‌های جدولی و کتابخانه‌های Matplotlib و Seaborn برای ترسیم نمودارهای توصیفی و بصری‌سازی استفاده شد. در گام نخست، یک ماتریس تراکنشی دودویی ساخته شد که در آن هر سطر نمایانگر یک بیمار و هر ستون نمایانگر یک داروی تجویزی بود؛ مقدار ۱ نشان‌دهنده وجود دارو در نسخه بیمار و مقدار صفر نشان‌دهنده عدم تجویز آن دارو بود. در این فرایند، تکرار یک دارو در یک نسخه (مثلاً ثبت چندباره همان دارو) حذف شد، زیرا هدف تحلیل، صرفاً حضور/عدم حضور دارو در ترکیب تجویزی بود نه تعداد دفعات ثبت در همان نسخه. به‌منظور افزایش دقت تحلیل، نام و کد داروها استانداردسازی شد؛ بدین صورت که تفاوت‌های ناشی از ثبت نام تجاری در برابر نام ژنریک، تفاوت‌های نوشتاری و کدهای داخلی سامانه، همگی به یک شناسه یکتای دارویی نگاشت شدند. داروهایی که در تعداد بسیار کمی از تراکنش‌ها ظاهر شده بودند (بسیار نادر)، با هدف کاهش نویز، جلوگیری از استخراج قواعد ناپایدار و کاهش سربار محاسباتی از مجموعه داده حذف شدند. پس از این مراحل، مجموعه داده نهایی به‌شکل یک ماتریس تراکنشی با ابعاد بالا آماده اجرای الگوریتم‌های کاوش قوانین انجمنی شد.

الگوریتم‌ها و شاخص‌های داده‌کاوی

برای شناسایی الگوهای هم‌تجویزی داروها، دو الگوریتم شناخته‌شده در حوزه کاوش قوانین انجمنی به‌کار گرفته شد: الگوریتم Apriori و الگوریتم FP-Growth. الگوریتم Apriori با بهره‌گیری از خاصیت بسته بودن مجموعه‌های پرتکرار، به‌صورت تکرارشونده اقدام به تولید و ارزیابی مجموعه اقلام کاندیدا می‌کند، در حالی که الگوریتم FP-Growth با ساخت یک درخت فشرده (FP-tree) مجموعه اقلام پرتکرار را بدون تولید گسترده کاندیدا استخراج می‌نماید. استفاده هم‌زمان از این دو روش، امکان مقایسه نتایج و افزایش اطمینان نسبت به پایداری الگوهای کشف‌شده را فراهم کرد. قوانین انجمنی به‌صورت $X \rightarrow Y$ تعریف شدند که در آن X مجموعه داروهای مقدم (پیش‌نیاز) و Y مجموعه داروهای متعاقب (نتیجه) را نشان می‌دهد که در نسخه‌های بیماران به‌طور هم‌زمان ظاهر می‌شوند. برای ارزیابی قوانین استخراج‌شده، از سه معیار استاندارد استفاده شد:

۱-پشتیبانی: نسبت تراکنش‌هایی که در آنها X و Y به‌طور هم‌زمان ظاهر شده‌اند؛ به‌عنوان شاخص فراوانی الگو در کل داده.

۲- اطمینان: احتمال شرطی وقوع Y در صورت وقوع X؛ به‌عنوان شاخص قدرت پیش‌بینی قانون.

۳- ضریب بالابری: نسبت فراوانی مشاهده‌شده هم‌وقوعی X و Y به فراوانی مورد انتظار در حالت استقلال؛ مقادیر بزرگ‌تر از ۱ نشان‌دهنده وابستگی مثبت و مقادیر نزدیک به ۱ نشان‌دهنده هم‌وقوعی تصادفی است.

آستانه‌های حداقلی برای Support، Confidence و Lift بر اساس ملاحظات آماری و کاربرد بالینی در محیط Python تعریف شد تا: ۱- از شناسایی ارتباطات کاذب ناشی از داروهای بسیار نادر جلوگیری شود، ۲- قوانینی که تنها در بخش بسیار کوچکی از جمعیت رخ می‌دهند و فاقد اهمیت جمعیتی هستند حذف شوند و ۳- نتایج به قوانینی محدود شود که از نظر فراوانی، قدرت پیش‌بینی و وابستگی، از پایداری و قابلیت تفسیر بالینی برخوردارند. تنها قوانینی که هر سه آستانه را به‌طور هم‌زمان برآورده می‌کردند، به‌عنوان قوانین معتبر پذیرفته شده و مبنای گزارش و تفسیر نتایج قرار گرفتند.

[DOI: 10.66224/ihio.9.1.6]

تحلیل آماری

در کنار تحلیل قوانین انجمنی، توصیف اولیه داده‌ها شامل محاسبه فراوانی مطلق و نسبی داروهای تجویزی (اختصاصی و غیراختصاصی)، میانگین و انحراف معیار تعداد دارو در هر نسخه انجام شد. در صورت نیاز، از آزمون‌های آماری توصیفی و نمودارهای هیستوگرام، جعبه‌ای و Heatmap برای نمایش الگوهای هم‌وقوعی و توزیع فراوانی استفاده گردید. همه تحلیل‌ها در محیط Python اجرا شد و سطح معناداری در تحلیل‌های تکمیلی (در صورت استفاده از آزمون‌های استنباطی) معادل ۰,۰۵ در نظر گرفته شد. این ساختار روش‌شناختی به گونه‌ای تدوین شده است که هر پژوهشگر دیگری، با دسترسی به داده‌های نسخه‌ای مشابه، بتواند مراحل پیش‌پردازش، مدلسازی تراکنشی، اجرای الگوریتم‌های Apriori و FP-Growth و اعمال آستانه‌های Support، Confidence و Lift را تکرار کرده و به نتایج قابل مقایسه دست یابد.

یافته‌ها

تحلیل توصیفی تجویز داروها در بیماران مبتلا به آلزایمر

پس از بارگذاری داده‌ها با استفاده از زبان برنامه‌نویسی پایتون، مجموعه‌داده نهایی شامل ۵۳۲٬۳۶۹ رکورد بیمار و ۶۴ متغیر بود که از میان آنها ۵۰ متغیر به داروهای تجویزی اختصاص داشتند. این حجم بالای داده امکان بررسی الگوهای تجویز در مقیاس بزرگ را فراهم کرد. نتایج تحلیل فراوانی داروها (جدول ۱) نشان داد که الگوی تجویز به‌شدت متمرکز بر داروهای استاندارد درمان آلزایمر است.

جدول ۱. لیست داروهای پرتکرار و فراوانی تجویز آنها

همانطور که از نتایج جدول ۱ قابل مشاهده است، Donepezil و Memantine با اختلاف قابل توجهی نسبت به سایر داروها بیشترین فراوانی را نشان دادند، در حالی که Galantamine سهم بسیار محدودی داشت. در مقابل، داروهای غیراختصاصی (مانند ویتامین‌ها، داروهای ضدالتهاب، قلبی–عروقی یا روان‌پزشکی) تنها در تعداد بسیار اندکی از موارد مشاهده شدند. این موضوع نشان می‌دهد که بخش عمده تجویزها با راهنماهای بالینی درمان آلزایمر همخوانی دارد و استفاده از داروهای غیرمرتبط محدود بوده است.

نتایج کاوش قوانین انجمنی

اجرای الگوریتم‌های Apriori و FP-Growth بر روی داده‌های تجویز دارو، که پس از پیش‌پردازش و تبدیل به قالب تراکنشی آماده شده بودند، برای کشف الگوهای هم‌تجویزی داروها انجام شد. داده‌ها ابتدا به ماتریس دودویی (Presence/Absence) تبدیل شدند و سپس شاخص‌های Support، Confidence و Lift برای تمام زوج‌های دارویی محاسبه گردید. نتایج تحلیل عددی نشان داد که تعداد کل قوانین بالقوه استخراج‌شده برابر با ۳۸۰ قانون بوده توزیع شاخص‌ها مطابق با جدول ۲ است.

این اعداد بیانگر آن هستند که بسیاری از قوانین کاندید، پشتیبانی بسیار پایین داشتند؛ به این معنا که ترکیبات دارویی مربوطه تنها در بخش کوچکی از بیماران مشاهده شده و فاقد اعتبار جمعیتی بودند. برخی از قوانین، مقادیر اطمینان نزدیک به احتمال پایه تجویز داروها داشتند؛ بنابراین قدرت پیش‌بینی محدود بود. اکثر قوانین، مقادیر

جدول ۱- لیست داروهای پرتکرار و فراوانی تجویز آنها

نام دارو	فراوانی تجویز
Donepezil	۲۲۰,۵۱۷
Memantine	۲۱۷,۴۱۰
Rivastigmine	۷۹,۶۵۳
Galantamine	۳,۹۶۳
سایر داروها	۲۲۱ و کمتر

جدول ۲- توزیع شاخص‌ها براساس مشاهدات

شاخص	count	mean	std	min	٪۲۵	٪۵۰	٪۷۵	max
Support	۳۸۰	$۵,۹۳\times۱۰^{-۸}$	$۴,۶۹\times۱۰^{-۷}$	۰	۰	۰	۰	$۵,۶۳\times۱۰^{-۶}$
Confidence	۳۸۰	۰,۰۰۷۹	۰,۰۶۸	۰	۰	۰	۰	۱

[DOI: 10.66224/ihio.9.1.6]

ضریب بالابری نزدیک یا کمتر از ۱ داشتند؛ بدین معنا که وابستگی مشاهده‌شده بین داروها تفاوت معناداری با وقوع همزمان تصادفی ندارد.

در مجموع، نتایج نشان می‌دهد که الگوریتم‌های کلاسیک کاوش قوانین انجمنی، نظیر Apriori و FP-Growth، به‌تنهایی قادر به شناسایی الگوهای معنادار تجویز داروهای مرتبط با آلزایمر در داده‌های واقعی بزرگ‌مقیاس نیستند و تحلیل‌های تکمیلی یا روش‌های پیشرفته‌تر، مانند استخراج الگوهای متوالی یا خوشه‌بندی طولی مسیرهای بیمار، ممکن است برای کشف الگوهای پیچیده‌تر ضروری باشند.

تفسیر نمودار پراکندگی

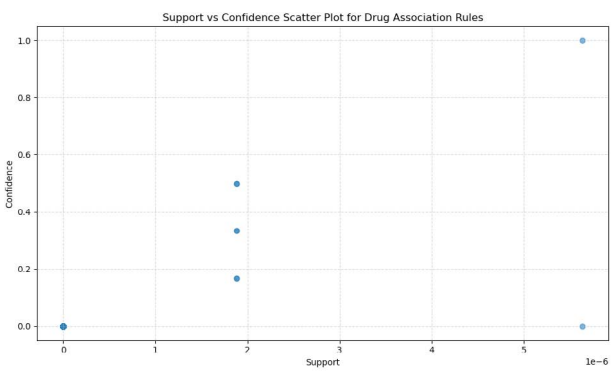
نمودار پراکندگی (شکل-۱) Support در مقابل Confidence برای قوانین استخراج‌شده، تصویر بصری روشنی از نتایج عددی ارائه می‌دهد. تمرکز اغلب نقاط در نزدیکی مبدأ مختصات نشان می‌دهد که اغلب قوانین دارای Support بسیار ناچیز هستند و حتی قوانینی با Confidence بالا (نزدیک به ۱) نیز Support تقریباً صفر دارند. این پدیده بیانگر آن است که برخی قوانین ممکن است در موارد بسیار نادری رخ داده باشند (مثلاً یک یا دو بیمار)، اما این هم‌وقوعی‌ها فاقد قابلیت تعمیم و ارزش بالینی هستند. بنابراین، مشاهده Confidence بالا به‌تنهایی نمی‌تواند به‌عنوان نشانه‌ای از وجود یک الگوی درمانی معتبر تلقی شود و باید همزمان با Support تفسیر گردد.

تحلیل Heatmap هم‌وقوعی داروها

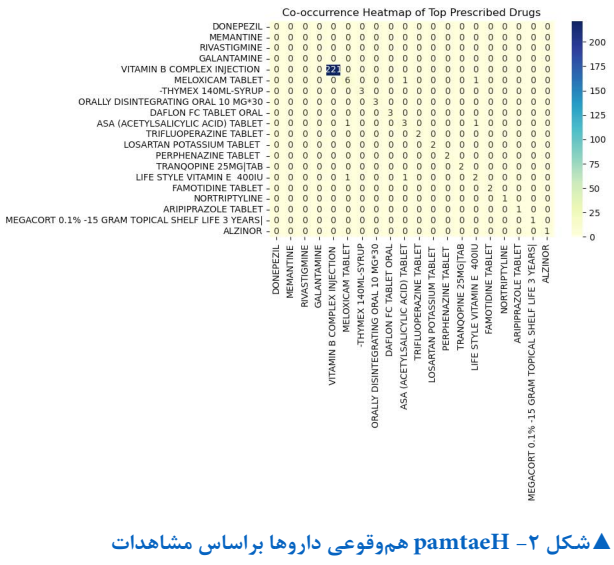
Heatmap هم‌وقوعی داروها، توزیع مشترک تجویز داروها بین بیماران را نشان می‌دهد و امکان شناسایی الگوهای بالقوه هم‌تجویزی یا گروه‌بندی داروهای معمول را فراهم می‌کند. در این مطالعه، از ماتریس دودویی داروها استفاده شد و تعداد بیمارانی که هر دو دارو را همزمان دریافت کرده‌اند، به‌عنوان مقدار هر خانه در ماتریس نمایش داده شد (شکل ۲).

نمودار Heatmap نشان می‌دهد که تراکم بسیار بالا در داروهای اصلی آلزایمر وجود داشته و Donepezil و Memantine بیشترین هم‌وقوعی را داشتند، اما حتی این مقادیر نسبت به حجم کل بیماران بسیار محدود هستند. Rivastigmine و Galantamine نسبت به Donepezil–Memantine، نقاط هم‌وقوعی کمتری دارند.

رنگ‌های روشن و نقاط پراکنده برای سایر داروها نشان می‌دهد که داروهای جانبی یا غیرمرتبط (ویتامین‌ها، داروهای قلبی–عروقی، ضدالتهاب و غیره) تقریباً هیچ هم‌وقوعی با داروهای اصلی یا با یکدیگر ندارند. این نشان



▲شکل ۱- نمودار پراکندگی براساس مشاهدات دارویی



▲شکل ۲- pamtaeH هم‌وقوعی داروها براساس مشاهدات

[DOI: 10.66224/ihio.9.1.6]

می‌دهد که این داروها در تعداد بسیار کمی از بیماران استفاده شده‌اند و هیچ خوشه قابل توجهی ایجاد نمی‌کنند.

همانطور که از نقشه رنگ ماتریس قابل مشاهده است، ماتریس تقریباً «رقیق» است و هیچ بلوک یا خوشه‌ی برجسته‌ای مشاهده نمی‌شود. این نتیجه به‌وضوح با تحلیل Rules و Scatter Plot Support–Confidence همخوانی دارد و نشان می‌دهد که هیچ مجموعه دارویی پرتکراری که به‌طور مداوم با یکدیگر تجویز شده باشد، وجود ندارد، یعنی الگوهای هم‌تجویزی پایداری برای این نوع داده وجود ندارد.

مقایسه بین الگوریتم‌ها

نتایج اجرای الگوریتم‌های Apriori و FP-Growth نشان داد که هر دو روش الگوهای هم‌وقوعی مشابهی استخراج کرده و از نظر محتوایی کاملاً قابل مقایسه هستند. الگوریتم FP-Growth به‌دلیل بهره‌گیری از ساختار درخت FP-tree و اجتناب از تولید گسترده مجموعه‌های کاندیدا، کارایی محاسباتی بالاتری در پردازش داده‌های بزرگ و پراکنده از خود نشان داد، اما پس از اعمال آستانه‌های حداقل Support، Confidence و Lift، هیچ‌یک از الگوریتم‌ها نتوانستند قوانین انجمنی با ارتباط بالینی معنادار تولید کنند. این همگرایی نتایج حاکی از آن است که عدم شناسایی قوانین معتبر، ناشی از ویژگی‌های ذاتی داده‌ها (پراکندگی بالا و پشتیبانی پایین) بوده و نه محدودیت الگوریتمی خاص؛ بنابراین، پایداری یافته‌ها مستقل از روش انتخابی تأیید شده و بر پیچیدگی الگوهای تجویزی واقعی در آلزایمر و ضرورت رویکردهای پیشرفته‌تر تأکید می‌نماید.

بحث

نتایج این مطالعه نشان داد که الگوریتم‌های Apriori و FP-Growth، علی‌رغم کارایی محاسباتی مناسب FP-Growth در داده‌های بزرگ، قادر به استخراج قوانین انجمنی با معیارهای Support، Confidence و Lift معنادار نشدند. عدم موفقیت الگوریتم‌های Apriori و FP-Growth در استخراج قوانین انجمنی معنادار با Support، Confidence و Lift مناسب، بیانگر پراکندگی بالای الگوهای تجویزی در داده‌های واقعی آلزایمر است. تمرکز شدید بر Donepezil و Memantine با هم‌وقوعی بالا، اما فقدان الگوهای پایدار در سایر ترکیبات، نشان‌دهنده سادگی نسبی الگوهای تجویزی در سطح جمعیتی است. این یافته با مطالعات داده‌کاوی پزشکی همخوانی دارد که در آنها، داده‌های تراکنشی بالینی اغلب Support پایین (کمتر از ۰/۰۰۱) نشان می‌دهند و الگوهای قوی انجمنی نادر هستند [۱۰].

در حالی که مطالعات غربی بر داده‌های بیمه‌ای کوچک‌تر، هم‌وقوعی Donepezil–Memantine را با Lift>۱,۲ گزارش کرده‌اند، داده‌های ایرانی با حجم عظیم (۵۳۲ هزار پرونده) تمرکز بالاتری بر ChEIs و NMDA antagonists را تأیید می‌کند که با راهنماهای NICE و AAN سازگار است [۱۱]. با این حال، عدم شناسایی قوانین معنادار حاکی از ناهمگونی بیشتر بالینی (هم‌ابتلایی‌ها، تفاوت‌های منطقه‌ای، ترجیحات تجویزی) نسبت به مطالعات قبلی است. مقایسه Apriori و FP-Growth نیز کارایی بالاتر FP-Growth را در داده‌های sparse تأیید می‌کند، مشابه گزارش‌های کاوش داروخانه‌ای [۱۲].

مدل تراکنشی دودویی (Presence/Absence) بدون در نظر گیری دوز، مدت درمان یا ترتیب زمانی، الگوهای Sequential را نادیده گرفت و Sequential Pattern Mining را ضروری ساخت. حذف داروهای نادر (Rare Items) اگرچه نویز را کاهش داد، ممکن است الگوهای بالینی خاص (مانند تجویزهای مرحله شدید) را حذف کرده باشد. ماهیت retrospective داده‌های Claims نیز سوگیری ثبت (under-diagnosis آلزایمر) و عدم دسترسی به شدت بالینی (MMSE/ADL) را به همراه دارد [۱۳]. با وجود این محدودیت‌ها، همگرایی نتایج دو الگوریتم اعتبار

[DOI: 10.66224/ihio.9.1.6]

یافته‌ها را تقویت می‌کند.

تمرکز تجویزی بر داروهای اثبات‌شده (Donepezil ۴۱ درصد،Memantine ۴۰ درصد) پایبندی پزشکان ایرانی به پروتکل‌های استاندارد را نشان می‌دهد، اما حضور داروهای غیرداختصاصی (SSRI ها، NSAID ها) پتانسیل پولی‌فارماسی را برجسته می‌سازد که ریسک عوارض Iatrogenic و هزینه‌ها را افزایش می‌دهد [۷]. با پیش‌بینی ۵۷ درصد رشد بیماران آلزایمر تا ۲۰۲۹، این یافته‌ها سیاست‌گذاران را به آموزش تجویزی مبتنی بر شواهد، محدودسازی داروهای کم‌ارتباط و پایش پولی‌فارماسی سوق می‌دهد [۷]. نتایج همچنین مبنایی برای CDS (Clinical Decision Support) Systems فراهم می‌کند

رویکردهای Sequential Rule Mining (مانند SPADE/ GSP) برای الگوهای زمانی، Graph Neural Networks برای روابط پیچیده دارو-بیمار-شدت و Ensemble Clustering بر اساس دموگرافیک‌ها مناسب‌ترند. ادغام EHR جامع با داده‌های بالینی (MMSE، neuroimaging) و ML predictive models تفسیر را بهبود می‌بخشد. مطالعات آینده می‌توانند نابرابری‌های منطقه‌ای/اجتماعی-اقتصادی را بررسی کرده و مدل‌های causal inference (مانند Propensity Score Matching) برای روابط علت-معلولی تجویز اعمال نمایند.

نتیجه‌گیری

این مطالعه بر روی داده‌های واقعی ۵۳۲٬۳۶۹ نسخه آلزایمر (۱۳۹۹-۱۴۰۱) نشان داد که الگوریتم‌های Apriori و FP-Growth قادر به استخراج قوانین انجمنی معنادار نشده و الگوهای تجویزی عمدتاً بر Donepezil و Memantine متمرکز است، که بیانگر پایبندی نسبی به راهنماهای بالینی اما پیچیدگی بالای داده‌های واقعی است. یافته کلیدی، محدودیت روش‌های کلاسیک در داده‌های Sparse بالینی و ضرورت رویکردهای Sequential/Graph-Based است؛ این پژوهش اولین تحلیل جامع قوانین انجمنی بر داده‌های بیمه سلامت ایران بوده و مبنایی برای بهینه‌سازی تجویز، کاهش پولی‌فارماسی و سیاست‌گذاری مبتنی بر شواهد در برابر ۵۷ درصد رشد بیماران آلزایمر تا ۲۰۲۹ فراهم می‌آورد.

تأییدیه اخلاقی: پروتکل مطالعه توسط کمیته اخلاق در پژوهش دانشگاه علوم پزشکی تهران بررسی و تأیید شده و این طرح با کد اخلاق IR.TUMS.NIHR.REC.1403.003 به ثبت رسیده است.

تضاد منافع: نویسندگان این مقاله اعلام داشتند که هی چگونه تعارض منافعی وجود ندارد.

سهم‌نویسندگان: شیما محمدی (۲۵٪): نویسنده مسئول؛ ایده‌پردازی و هماهنگی مطالعه، دریافت و هماهنگی دسترسی به داده‌های سازمان بیمه سلامت ایران، نگارش و ویرایش مقاله، تفسیر یافته‌ها و نظارت کلی بر مطالعه؛ مائده غلام‌آزاد (۳۰٪): تحلیل داده‌ها، رویه‌های داده‌کاوی، پیاده‌سازی الگوریتم‌های کاوش قوانین انجمنی (Apriori و FP-Growth)، تفسیر نتایج، بررسی و ویرایش مقاله؛ علیرضا آتشی (۲۰٪): پردازش و دسته‌بندی داده‌ها، تحلیل‌های آماری و محاسباتی، بررسی و ویرایش علمی مقاله؛ فاطمه حاجی‌علیاسگری (۱۰٪): آماده‌سازی داده‌ها، سازماندهی و طبقه‌بندی داده‌ها، همکاری در دریافت و اعتبارسنجی داده‌ها از سازمان بیمه سلامت؛ لیلا غمخوار (۱۰٪): آماده‌سازی داده‌ها، سازماندهی و طبقه‌بندی داده‌ها، همکاری در جمع‌آوری و مدیریت داده‌ها از سازمان بیمه سلامت؛ رضا مهریزی (۵٪): شناسایی و طبقه‌بندی داروهای مرتبط با بیماری آلزایمر، مشاوره بالینی و مشارکت در تهیه مقاله؛ حسین عفت‌پناه (۵٪): مشاوره علمی، تفسیر یافته‌ها و بررسی مقاله.

منابع مالی: پژوهش حاضر از حمایت مالی سازمان بیمه سلامت ایران برخوردار بوده است.

References

1. Mohammadi S, Zandi M, Dousti Kataj P, Karimi Zandi L. Chronic stress and Alzheimer's disease. Biotechnol Appl Biochem. 2022;69(4):1451-8. doi: 10.1002/bab.2216.
2. Iulita MF, Garzón Chavez D, Klitgaard Christensen M, Valle Tamayo N, Plana-Ripoll O, Rasmussen SA, et al. Association of Alzheimer Disease With Life Expectancy in People With Down Syndrome. JAMA Netw Open. 2022;5(5):e2212910. doi: 10.1001/jamanetworkopen.2022.12910.
3. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement. 2013;9(1):63-75.e2. doi: 10.1016/j.jalz.2012.11.007.
4. Borhaninejad V, Gohari MA, Jahani Y, Googhari SA, Jannati N, Mirzaee M. Population Forecasting with Alzheimer's Disease In Iran Using a System Dynamic Model. Iran J Public Health. 2025;54(5):1064-73. doi: 10.18502/ijph.v54i5.18641.
5. Zhang J, Zhang Y, Wang J, Xia Y, Zhang J, Chen L. Recent advances in Alzheimer's disease: Mechanisms, clinical trials and new drug development strategies.

[DOI: 10.66224/ihio.9.1.6]

Signal Transduct Target Ther. 2024;9(1):211. doi: 10.1038/s41392-024-01911-3.
6. Olivares D, Deshpande VK, Shi Y, Lahiri DK, Greig NH, Rogers JT, et al. N-methyl D-aspartate (NMDA) receptor antagonists and memantine treatment for Alzheimer's disease, vascular dementia and Parkinson's disease. Curr Alzheimer Res. 2012;9(6):746-58. doi: 10.2174/156720512801322564.
7. Ataei SM, Golestani A, Khosravi S, Tabatabaei-Malazy O, Malekpour MR, Ebrahimpur M, et al. Medication use patterns and polypharmacy among elderly in Iran: a cross-sectional study using national health insurance claims data. BMJ Open. 2025;15(8):e097863. doi: 10.1136/bmjopen-2024-097863.
8. El Khalifa RA, Yew PY, Chi CL. Detecting Multimorbidity Patterns with Association Rule Mining in Patients with Alzheimer's Disease and Related Dementias. AMIA Jt Summits Transl Sci Proc. 2024;2024:525-34.
9. Moghaddassi H, Hoseini A, Asadi F, Jahanbakhsh M. Application of data mining in health. Health Information Management. 2012;9(2):297-304. [persian]
10. Ridwan LJSD. Comparative Analysis of Apriori, FP-Growth, and ECLAT for Mining Medical Association Patterns; 2025.
11. Pujiharto EW, Kusrini K, Nasiri AJCSJ. Analisis Perbandingan Kinerja Algoritma Apriori, FP-Growth dan Eclat dalam menemukan Pola Frekuensi pada Dataset INA-CBG'S. 2023;9(2):340-54. doi: 10.31154/cogito.v9i2.547.340-354.
12. Husain Y, Oktaviyani ED, Christina S. Analisis perbandingan algoritma Apriori, FP-Growth, dan Eclat dalam menemukan pola pembelian konsumen. KONSTELASI: Konvergensi Teknologi dan Sistem Informasi. 2023;3(2):231-43. doi: 10.24002/konstelasi.v3i2.7007.
13. Spooner A, Mohammadi G, Sachdev PS, Brodaty H, Sowmya A. Identifying risk factors for Alzheimer's disease from multivariate longitudinal clinical data using temporal pattern mining. BMC Bioinformatics. 2025;26(1):56. doi: 10.1186/s12859-024-06018-8.

[DOI: 10.66224/ihio.9.1.6]