DOR: 20.1001.1.2538161.1404.10.4.5.7

DOI : 10.22034/csj.2025.237376

زمستان ۱۴۰۴، دوره ۱۰، شماره ۴، پیاپی ۳۹

دریافت مقاله: ۱۴۰۴/۰۳/۲۷

پذیرش مقاله: ۱۴۰۴/۰۶/۲۳

نوع مقاله: پژوهشی

خوشه‌بندی بیماران مبتلا به مولتیپل اسکلروزیس در استان چهارمحال و بختیاری با استفاده از الگوریتم K-prototypes فازی و غیرفازی

وحیده احراری*

استادیار گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه شهرکرد، ایران

پست الکترونیکی: vahideh.ahrari@sku.ac.ir

زهرا بهدانی

استادیار گروه آمار و ریاضی، دانشکده انرژی و علوم داده، دانشگاه صنعتی خاتم الانبیاء بهبهان، ایران

پست الکترونیکی: behdani@bkatu.ac.ir

چکیده

خوشه‌بندی به‌عنوان یکی از روش‌های یادگیری بدون ناظر، ابزاری مؤثر برای کشف الگوهای پنهان و ساختارهای معنادار در داده‌هاست. در حوزه سلامت، این روش امکان تفکیک بیماران بر اساس ویژگی‌های بالینی و درمانی را فراهم می‌کند. مطالعه حاضر با بهره‌گیری از الگوریتم K-prototypes فازی و غیرفازی، به شناسایی زیرگروه‌های بالینی بیماران ام. اس با هدف بهبود برنامه‌ریزی درمانی و کیفیت زندگی آنان می‌پردازد. در این تحقیق، اطلاعات مربوط به ٤٤٨ بیمار مبتلا به مولتیپل اسکلروزیس، شامل ویژگی‌های جمعیت‌شناختی و نوع داروهای مصرفی در استان چهارمحال و بختیاری استفاده شد. جهت شناسایی الگوهای نهفته در داده‌ها، از الگوریتم خوشه‌بندی K-prototypes در دو حالت فازی و غیرفازی استفاده گردید. تعیین تعداد بهینه خوشه‌ها با بهره‌گیری از شاخص‌های سیلوئت، دیویس–بولدین و کالینسکی–هاراباسز انجام شد. نتایج تحلیل نشان داد که سه خوشه به عنوان بهترین تفکیک برای داده‌ها انتخاب شدند. بررسی شاخص‌های ارزیابی نشان داد که عملکرد دو الگوریتم مورد استفاده نزدیک به هم بوده، ولی الگوریتم فازی عملکرد نسبتاً بهتری از خود نشان داده است. توصیف کامل خوشه‌ها در بخش نتایج به تفصیل بیان شده است. این مطالعه نشان می‌دهد که روش‌های خوشه‌بندی فازی می‌توانند در تحلیل الگوهای پنهان بیماران ام.اس مؤثر باشند و در طراحی مداخلات درمانی مبتنی بر ویژگی‌های فردی به‌کار روند. این رویکرد نه تنها قابلیت ارائه بینش‌های جدید در مدیریت بیماری‌های مزمن را فراهم می‌سازد، بلکه می‌تواند در شناخت الگوهای بیماری و برنامه‌ریزی درمانی مؤثر باشد. یافته‌های این پژوهش می‌توانند مبنای علمی ارزشمندی برای تصمیم‌گیری‌های بالینی فراهم کنند

واژه‌های کلیدی: یادگیری ماشین، خوشه‌بندی، خوشه‌بندی فازی، خوشه‌بندی K-prototypes و مولتیپل اسکلروزیس.

۱- مقدمه

مولتیپل اسکلروزیس[1] (ام.اس) یک اختلال خودایمنی

1-Multiple Sclerosis

* نویسندهٔ مسئول

مزمن است که بر سیستم عصبی مرکزی از جمله مغز، نخاع و اعصاب بینایی تأثیر می‌گذارد و باعث اختلال در انتقال سیگنال‌های عصبی می‌شود. به طور معمول این بیماری در بزرگسالان و زنان جوان شیوع بیشتری دارد. اگرچه عوامل ژنتیکی، محیطی و ... در این بیماری دخیل هستند، اما با وجود تحقیقات گسترده، علت دقیق این بیماری ناشناخته باقی‌مانده است. ناهمگونی این بیماری چالش‌هایی را در زمینه‌های مختلف مرتبط با این بیماری ایجاد کرده است. در حال حاضر، رویکردهای درمانی بر مدیریت علائم، کاهش فعالیت بیماری و اصلاح دوره بیماری با هدف بهبود کیفیت زندگی بیماران متمرکز است [۱].

در سال‌های اخیر، تحلیل داده در مراقبت‌های سلامت به‌ویژه روش‌های یادگیری بدون ناظر به ابزاری مؤثر برای کشف الگوهای پنهان و پشتیبانی از تصمیم‌سازی بالینی بدل شده است. خوشه‌بندی به‌عنوان هستهٔ این رویکردها با افراز داده‌ها به گروه‌های درون‌همگن و بین‌ناهمگن، نمونه‌های مشابه را گرد هم می‌آورد و تفسیر داده‌ها را ساده‌تر می‌کند. رویکردهای آن به‌طورکلی به دو دسته سلسله‌مراتبی (تولید دندروگرام بر پایهٔ معیار شباهت) و افرازی (ساخت k خوشه از پیش تعیین شده) تقسیم می‌شوند [۲، ۳].

در روش‌های خوشه‌بندی غیرسلسله مراتبی، تعداد k خوشه با حداقل یک عضو ساخته می‌شود و این الگوریتم‌ها به دو دسته سخت (انحصاری) و نرم (فازی) تقسیم می‌شوند. در روش‌های خوشه‌بندی کلاسیک دو شرط کلی برقرار است: هر خوشه باید حداقل دارای یک عضو باشد و هر عضو باید فقط به یک خوشه متعلق باشد. تفاوت اصلی بین خوشه‌بندی کلاسیک و فازی این است که در خوشه‌بندی فازی، شرط انحصاری بودن نادیده گرفته می‌شود و یک نمونه می‌تواند به چندین خوشه تعلق داشته باشد. الگوریتم خوشه‌بندی مناسب، براساس هدف کاربر از تحلیل داده و نوع داده‌های موجود انتخاب شود. تا سال ۱۹۶۵ روش‌های خوشه‌بندی به صورت قطعی استفاده می‌شدند. بنیان نظری خوشه‌بندی فازی بر مجموعه‌های فازی زاده استوار است و در کارهای بعدی گسترش یافته است [۷- ٤]. در عمل، الگوریتم‌های خانوادهٔ C- میانگین و نسخهٔ فازی آن، FCM به‌سبب سادگی و همگرایی مناسب، کاربرد گسترده‌ای دارند، اما نسبت دادن وزن یکنواخت به ویژگی‌ها و حساسیت به فاصلهٔ اقلیدسی در حضور نوفه/ پرت از محدودیت‌های رایج آن‌هاست. از این‌رو، یادگیری وزن ویژگی‌ها و استفاده از فواصل غیراقلیدسی برای بهبود پایداری و دقت پیشنهاد شده است [۱٤- ۷].

برانکو و همکاران [۱۵] از روش‌های مختلف یادگیری ماشین، از جمله جنگل‌های تصادفی برای توسعه مدل‌های پیش‌بینی‌کننده برای پیشرفت و نتایج درمان بیماری ام.اس استفاده کردند. اسحاقی و همکاران [۱٦] از روش‌های یادگیری بدون ناظر با داشتن تصویربرداری تشدید مغناطیسی بیماران ام.اس استفاده کرده و بیماران را بر اساس شباهت ویژگی‌هایشان خوشه‌بندی کردند. خوشه‌های به دست آمده براساس ویژگی‌هایی از جمله شدت علائم، سرعت پیشرفت و پاسخ به درمان تجزیه‌وتحلیل شدند. این تجزیه‌وتحلیل، به شناسایی الگوهای متمایز بیماران ام.اس کمک کرده و می‌تواند راهبردهای درمانی داشته باشد.

این پژوهش با تمرکز بر تحلیل خوشه‌ای بیماران مبتلا به ام.اس در استان چهارمحال و بختیاری انجام شده است. مجموعه داده شامل ٤٤۸ بیمار بوده و چهار ویژگی کلیدی جمعیت‌شناختی و دارویی شامل سن، جنس، سن در زمان تشخیص بیماری و نوع داروی مصرفی را در بر می‌گیرد. با وجود پیشرفت‌های روش‌شناختی، ادبیات موجود با چند محدودیت اساسی مواجه است: نخست، بسیاری از رویکردها داده‌های ترکیبی (عددی- طبقه‌ای) را به‌صورت یکپارچه و با وزن‌دهی پویا به ویژگی‌ها پشتیبانی نمی‌کنند، به‌عنوان مثال، نسخه‌های پایه K-modes و K-prototypes

غالباً از وزن‌های یکنواخت یا معیارهای سادهٔ عدم‌تشابه استفاده می‌کنند [۱۸، ۱۷]. دوم، شماری از مطالعات بالینی تنها بر یک بُعد داده مثلاً MRI متمرکز بوده و متغیرهای دارویی یا جمعیت‌شناختی را نادیده گرفته‌اند که این امر تفسیر بالینی را محدود می‌سازد [۱۶]. سوم، ارزیابی کیفیت خوشه‌ها اغلب به یک شاخص منفرد متکی است و کمتر به سنجش پایداری ساختار توجه شده است [۱۵].

برای پاسخ به این شکاف‌ها، در این مطالعه نسخه‌ای فازی از الگوریتم K-prototypes با قابلیت وزن‌دهی تطبیقی به ویژگی‌ها و معیار عدم‌تشابه سازگار با داده‌های ترکیبی ارائه شد. الگوریتم پیشنهادی بر روی مجموعه‌داده واقعی بیماران استان چهارمحال‌وبختیاری ارزیابی گردید و کیفیت و بهینگی خوشه‌ها با سه شاخص معتبر سیلوئت، دیویس- بولدین و کالینسکی- هاراباسز سنجیده شد. این پژوهش درصدد پاسخ به سه پرسش اصلی است: ۱. آیا می‌توان زیرگروه‌های متمایز بیماران اماس را بر اساس ویژگی‌های جمعیت‌شناختی و دارویی شناسایی کرد؟ ۲. عملکرد روش‌های فازی و غیرفازی K-prototypes در خوشه‌بندی این داده‌ها چگونه است؟ ۳. خوشه‌های شناسایی‌شده چگونه می‌توانند در بهبود مدیریت بالینی و شخصی‌سازی درمان بیماران ام.اس مؤثر باشند؟

مهم‌ترین دستاوردهای پژوهش را می‌توان در سه محور اصلی خلاصه کرد: اول، ارائه و ارزیابی الگوریتم K- prototypes فازی با وزن‌دهی تطبیقی به ویژگی‌ها که نسبت به الگوریتم پایه، باعث افزایش دقت خوشه‌بندی شد. دوم، استفاده هم‌زمان از سه معیار ارزیابی خوشه‌بندی برای تعیین تعداد بهینه خوشه‌ها در داده‌های پزشکی و سوم، کشف الگوهای درمانی خاص (مانند مصرف غالب ربیف/ بتافرون در بیماران جوان در مقابل آونکس در سالمندان) که به‌عنوان راهنمای بالینی برای تجویز داروی شخص‌محور و مدیریت سیر بیماری عمل می‌کند. در ادامه، بررسی کارهای مرتبط (بخش ۲)، مبانی نظری الگوریتم‌ها (بخش‌های ۵-۳)، روش‌شناسی (بخش ٦)، یافته‌ها (بخش ۷) و بحث و نتیجه‌گیری (بخش ۸) ارائه می‌گردد.

۲- تحقیقات مرتبط

مطالعات پیشین در حوزه خوشه‌بندی بیماران مبتلا به ام.اس را می‌توان به دو دسته اصلی تقسیم کرد: مطالعات الگوریتمی و مطالعات بالینی. در دسته مطالعات الگوریتمی، به عنوان نمونه، هوانگ [۱۷] الگوریتم K-modes را معرفی کرد که گامی اساسی در خوشه‌بندی داده‌های طبقه‌ای محسوب می‌شود، ولی محدودیت‌هایی همچون عدم توانایی در پردازش هم‌زمان داده‌های عددی و طبقه‌ای، اختصاص وزن یکسان به همه ویژگی‌ها بدون توجه به اهمیت بالینی متفاوت آنها و استفاده از معیار فاصله ساده دارد. جی و همکاران [۱۸] به منظور غلبه بر مشکل داده‌های ترکیبی، نسخه فازی الگوریتم K-prototypes را ارائه نمودند، اما در آن روش سازوکار وزن‌دهی پویا لحاظ نشده بود که می‌تواند منجر به کاهش دقت در تفکیک زیرگروه‌های ناهمگن گردد.

در حوزه بالینی، رویکردهای متنوعی به‌کار گرفته شده است. اسحاقی و همکاران [۱٦] با استفاده از داده‌های تصویربرداری MRI توانستند زیرگروه‌های متمایز از نظر ویژگی‌های عصبی را شناسایی کنند، اما به دلیل عدم بهره‌گیری از داده‌های دارویی و جمعیت‌شناختی (مانند سن در زمان تشخیص، جنسیت و الگوی مصرف دارو)، قابلیت کاربرد بالینی نتایج محدود ماند. برانکو و همکاران [۱۵] با استفاده از مدل جنگل تصادفی، پیش‌بینی سیر پیشرفت بیماری را هدف قرار دادند، ولی خروجی مطالعه فاقد استخراج الگوهای درمانی قابل تعمیم بود. همچنین، برجوئی و همکاران [۱۹] اگرچه مجموعه داده‌ای بومی و ارزشمند از جمعیت ایرانی را مورد استفاده قرار دادند، به‌دلیل تمرکز صرف بر عوامل ژنتیکی، تحلیل یکپارچه متغیرهای دارویی و جمعیت‌شناختی را در طراحی مدل خود لحاظ نکردند.

بررسی ادبیات موجود نشان می‌دهد که بخش

قابل‌توجـــهی از مطالعات خوشـــه‌بندی بیمـــاران مبتلا به مولتیپل اسکلروزیس، با وجود پیشرفت‌های روش‌شناختی، همچنان با محدودیت‌هـــایی در جنبه‌های کلیدی مواجه‌اند. بسیـــاری از رویکردهای الگوریتمی فاقد ســـازوکار مؤثر برای وزن‌دهی پویا به ویژگی‌ها هستند و این امر می‌تواند دقـــت و تفسیرپذیری خوشـــه‌ها را تحت تأثیـــر قرار دهد. همچنین، در بخشی از پژوهش‌ها، داده‌های مورد اســـتفاده تنها برخی جنبه‌های محدود از اطلاعات بیماران را پوشش داده و بنابرایـــن ارزیـــابی هم‌زمان چند ویـــژگی کلیدی با محدودیت همراه بوده اســـت. از نظر ارزیابی نیز، تکیه بر یک شـــاخص منفرد در بسیاری از مطالعـــات، پایداری و قابلیـــت اعتماد نتایج را کاهش می‌دهد. افزون بر این، پیوند دادن نتایج حاصل از خوشـــه‌بندی با کاربردهای عملی در تصمیم‌گیری بالینی کمتر مورد توجه قرار گرفته است.

در پاسخ به این محدودیت‌ها، پژوهش حاضر چارچوبی مبتنی بر الگوریتم K-prototypes فازی توسعه داده است که در آن، سازوکار وزن‌دهی تطبیقی به ویژگی‌ها پیاده‌سازی شـــده و تحلیل بر پایـــه چهار ویژگی جمعیت‌شـــناختی و دارویی شـــامل سن، جنسیت، سن تشخیص بیماری و نوع داروی مصرفی انجام گرفته است. ارزیابی کیفیت و بهینگی خوشـــه‌ها با بهره‌گیری هم‌زمان از ســـه شـــاخص معتبر سیلوئت، دیویس- بولدین و کالینسکی- هاراباسز صورت گرفته تا دقت و پایداری ســـاختار خوشه‌ها بررسی شود. این رویکرد امکان شناسایی الگوهای متمایز میان بیماران را فراهـــم کرده و می‌تواند در بهبـــود راهبردهای درمانی متناسب با ویژگی‌های فردی بیماران نقش داشته باشد.

مروری بر مهم‌ترین پژوهش‌های پیشین و تحلیل نقاط قوت و ضعف آن‌ها در جدول ۱ ارائه شده است که به‌طور خلاصه مقایســـه‌ای میان رویکردها و معیارهای ارزیابی به‌کاررفته در این مطالعات را نشان می‌دهد.

جدول ۱: مرور و مقایسه مطالعات منتخب در حوزه خوشه‌بندی بیماران مبتلا به مولتیپل اسکلروزیس

معیار ارزیابی	محدودیت‌ها و نقاط ضعف	نقاط قوت	روش‌شناسی	مطالعه
دیویس- بولدین	عدم توانایی در پردازش هم‌زمان داده‌های عددی و طبقه‌ای، وزن‌دهی یکنواخت به ویژگی‌ها، استفاده از معیار فاصله ساده	پردازش داده‌های طبقه‌ای	K-modes	Huang (۱۹۹۸)
تابع هدف فازی	عدم استفاده از سازوکار وزن‌دهی پویا برای ویژگی‌ها	مدیریت داده‌های ترکیبی	K-proto-types فازی	Ji et al (۲۰۱۲)
سیلوئت	عدم به‌کارگیری داده‌های دارویی و جمعیت‌شناختی	شناسایی زیرگروه‌های عصبی	خوشه‌بندی مبتنی بر MRI	Eshaghi (۲۰۲۱)
دقت	تمرکز صرف بر پیش‌بینی یک بعد از پیامدهای بیماری، فقدان استخراج الگوهای درمانی	پیش‌بینی پیشرفت بیماری	جنگل‌های تصادفی	Branco (۲۰۲۲)
آزمون‌های آماری	عدم تحلیل داده‌های مرتبط با الگوهای درمانی و جمعیت‌شناختی	داده‌های بومی ایرانی	تحلیل ژنتیکی	Barjui (۲۰۱۷)

به طور گســـترده در حوزه‌های مختلف از جمله تشخیص ناهنجاری، تجزیه‌وتحلیل تصویر، فشرده‌ســـازی داده‌ها و ... به کار گرفته شده‌اند. در پژوهش حاضر نیز از الگوریتم خوشـــه‌بندی K-prototypes استفاده شده که در ادامه به طور خلاصه به توضیح این الگوریتم‌ها پرداخته می‌شود.

الگوریتم k-means یک روش خوشه‌بندی است که برای تقسیم یک مجموعه داده با متغیرهای عددی به K خوشـــهٔ مجزا استفاده می‌شـــود. این الگوریتم با ایجاد خوشه‌های اولیه آغاز می‌شـــود و به دنبال آن نقاط داده به نزدیکترین مرکز خوشه تخصیص داده می‌شـــوند. پس از آن، مراکز خوشه بر اســـاس نقاط اختصاص داده شده جدید به روز می‌شوند. این چرخه تخصیص نقاط و به‌روزرسانی مراکز، تا زمانی ادامه می‌یابد که تخصیص‌های خوشه ثابت بماند [۳].

الگوریتم k-modes روش خوشـــه‌بندی است که برای مجموعه داده‌هایی با متغیرهای طبقه‌ای [۲] استفاده می‌شود. برخلاف الگوریتم k-means که بر مرکزهای عددی (مقادیر

۳- الگوریتم خوشه‌بندی K-prototypes

الگوریتم‌های خوشه‌بندی، مانند k-means و k-modes،

2-Categorical variables

میانگین) متکی است، k-modes از نما (مد) در هر خوشه استفاده میکند که مقادیر با بیشترین فراوانی برای هر متغیر طبقه‌ای در یک خوشه است [۱۷]. الگوریتم K-proto-types ترکیبی از الگوریتم‌های k-means و k-modes است که برای مجموعه داده‌هایی که شامل متغیرهای عددی و طبقه‌ای هستند، به کار می‌رود. این الگوریتم از نظر عملی مفیدتر از دو الگوریتم مزبور است، زیرا داده‌هایی که در حالت واقعی جمع‌آوری می‌شوند، معمولاً شامل هر دو نوع متغیر هستند.

یک مجموعه با n شیء را به صورت $X = \{X_1, X_2, \cdots, X_n\}$ در نظر بگیرید. فرض کنید $X_i = \{X_{i1}, X_{i2}, \cdots, X_{im}\}$ شامل m متغیر (که p متغیر آن عددی و $m - p$ متغیر آن طبقه‌ای هستند) باشد. هدف، خوشه‌بندی n شیء مذکور به k خوشه متمایز با مراکز خوشه $C = \{C_1, C_2, \cdots, C_k\}$ است. معیار عدم تشابه مورد استفاده در الگوریتم K-prototypes ترکیبی از فاصله اقلیدسی برای متغیرهای عددی و عدم تشابه ساده برای متغیرهای طبقه‌ای است. در این صورت فاصله $d(X_i, C_j)$ بین X_i و C_j را می‌توان به صورت زیر به دست آورد:

$$d(X_i, C_j) = d_r(X_i, C_j) + \gamma d_c(X_i, C_j), \qquad (۱)$$

که $d_r(X_i, C_j)$ فاصله بین متغیرهای عددی و $d_c(X_i, C_j)$ فاصله بین متغیرهای طبقه‌ای است و γ وزنی برای متغیرهای طبقه‌ای در نظر گرفته شده است، لذا داریم:

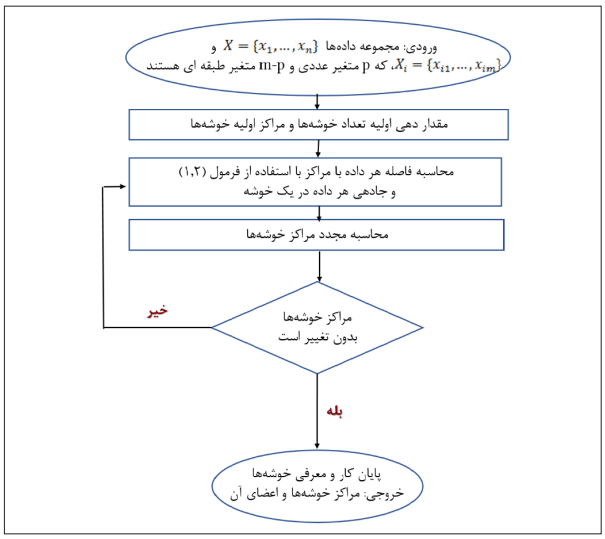
$$d_r(X_i, C_j) = \sum_{l=1}^{p} |x_{il} - c_{jl}|^2 \qquad (۲)$$

$$d_c(X_i, C_j) = \sum_{l=p+1}^{m} \delta(x_{il}, c_{jl}), \qquad (۳)$$

که در آن:

$$\delta(x_{il}, c_{jl}) = \begin{cases} 0, & x_{il} = c_{jl} \\ 1, & x_{il} \neq c_{jl}' \end{cases}$$

رابطه (۲)، مجذور فاصله اقلیدسی بین مراکز خوشه و مقادیر متغیرهای عددی را محاسبه می‌کند و رابطه (۳)، معیار عدم تشابه ساده برای داده‌های طبقه‌ای است. الگوریتم K-prototypes به طور مکرر نقاط داده را به خوشه با کمترین تفاوت اختصاص می‌دهد و نمونه‌های اولیه خوشه را تا زمان همگرایی به‌روز می‌کند. در شکل ۱ به صورت خلاصه چگونگی انجام این روش آمده است.



شکل ۱: الگوریتم خوشه‌بندی K-prototypes

۴- الگوریتم خوشه‌بندی K-prototypes فازی

خوشه‌بندی فازی، بخشی از تحلیل فازی است که از دو دیدگاه دارای اهمیت است. یکی تحلیل داده‌های فازی و دیگری تحلیل داده‌های قطعی با روش‌های فازی. در این مقاله، تاکید ما بر دیدگاه دوم است. ایده اصلی در خوشه‌بندی فازی این است که هر خوشه مجموعه‌ای از عناصر است و سپس با تغییر در تعریف عضویت عناصر در مجموعه کلاسیک، برای هر خوشه یک زیرمجموعه فازی در نظر گرفته می‌شود. لذا هر داده با یک درجه عضویت به هر یک از خوشه‌ها تعلق دارد.

یکی از اولین روش‌های خوشه‌بندی فازی در سال ۱۹۸۱، توسط بزدک [۲۰] معرفی شده است. این الگوریتم با عنوان C میانگین فازی بر مبنای تابع هدف فازی و استفاده از فاصله اقلیدسی معرفی شده است. الگوریتم C میانگین فازی به علت تعیین تصادفی پارامترهای اولیه و

از طرفی حساسیت آن به مقادیر تعیین شده، در هر اجرا خوشه‌های متفاوتی تولید می‌کند. همچنین برای شناسایی خوشه‌های متفاوت می‌توان فاصله اقلیدسی را با معیار دیگری جایگزین کرد. برای مثال گوستافسون و کسل [۲۱] از فاصله ماهالانبیس استفاده کردند که منجر به تولید خوشه‌های بیضی شکل شد.

جی و همکاران [۱۸] در سال ۲۰۱۲، الگوریتم خوشه‌بندی فازی K-prototypes را معرفی کردند. قبل از بیان الگوریتم و روش خوشه‌بندی مزبور، به ارائه دو تعریف زیر که در ادامه مورد نیاز است، می‌پردازیم.

تعریف ۱- برای یک مجموعه داده با m ویژگی، شامل ویژگی‌های عددی و طبقه‌ای که گسسته شده‌اند، فاصله بین دو مقدار x و y از هر ویژگی طبقه‌ای A_i به صورت زیر است:

$$\delta(x,y)=\frac{\sum_{j=1,i\neq j}^{m}\delta^{ij}(x,y)}{m-1} \qquad (4)$$

$\delta(x,y)$ در رابطه بالا دارای ویژگی‌های زیر است:

۱. $0\leq\delta(x,y)\leq 1,$

۲. $\delta(x,y)=\delta(y,x),$

۳. $\delta(x,x)=0$

اهمیت [۳] یک ویژگی، نقش آن ویژگی را در مجموعه داده‌ها مشخص می‌کند. آن دسته از ویژگی‌هایی که جداسازی خوبی روی مقادیر در گروه‌های مختلف را نشان می‌دهند، نقش مهمتری را در خوشه‌بندی اشیاء داده ایفا می‌کنند [۱۸]. به عبارت دیگر، یک ویژگی نقش مهمی در خوشه‌بندی ایفا می‌کند، مشروط بر این که هر جفتی از مقادیر ویژگی آن به خوبی از همه ویژگی‌های دیگر جدا شده باشد، به عنوان مثال، مقدار بالایی $d(x,y)$ برای همه جفت‌های x و y داشته باشد. بنابراین فاصله بین دو مقدار طبقه‌ای بیانگر اهمیت ویژگی مربوطه در آن مجموعه داده است.

3-Significance

برای ویژگی‌های عددی با مقیاس یکسان، فاصله بین دو مقدار عددی x و y برای یک ویژگی A_i با $d(x,y)=(w_i(x-y))^2$ محاسبه می‌شود، که در آن w_i نشان‌دهنده اهمیت ویژگی عددی است. برای محاسبه اهمیت یک ویژگی عددی، ابتدا ویژگی عددی را گسسته می‌کنیم. تعداد بازه‌های T را برای همه ویژگی‌های عددی انتخاب می‌کنیم. سپس به هر بازه یک مقدار طبقه‌بندی شده $u[1],u[2],\cdots,u[T]$، اختصاص داده می‌شود. پس از آن، برای ویژگی عددی گسسته، برای هر جفت مقادیر طبقه‌بندی $\delta(u[r],u[s])$ به همان شکلی که برای مقادیر طبقه‌ای استفاده می‌شود، فاصله را محاسبه می‌کنیم. سپس، اهمیت یک ویژگی عددی به عنوان میانگین همه جفت‌های $\delta(u[r],u[s])$ محاسبه می‌شود.

تعریف ۲- اهمیت w_i یک ویژگی عددی A_i به صورت زیر محاسبه می‌شود:

$$w_i=\frac{2\sum_{k=1}^{T}\sum_{j>k}^{T}\delta(u[k],u[j])}{T(T-1)} \qquad (5)$$

که در آن T تعداد بازه‌های گرفته شده است.

در ادامه الگوریتم خوشه‌بندی K-prototypes فازی را برای داده‌های ترکیبی ارائه می‌دهیم. فرض کنید X_i یک بردار از داده‌ها به صورت زیر است:

$$X_i=(x_{i,1}^r,x_{i,2}^r,\cdots,x_{i,p}^r,x_{i,p+1}^c,x_{i,p+2}^c,\cdots,x_{i,m}^c)$$

است که در آن r مولفه اول مقادیر عددی و بقیه مولفه‌ها مقادیر طبقه‌ای هستند. بنابراین prototypes دو بخش دارد. p مرکز اول، میانگین را برای ویژگی‌های عددی استفاده می‌کنند و بقیه مراکز، مراکز فازی داده‌های طبقه‌ای را استفاده خواهند کرد. تابع هدف در این روش خوشه‌بندی به صورت زیر تعریف می‌شود:

$$E(U,Q)=\sum_{j=1}^{k}\sum_{i=1}^{n}u_{ij}^{\alpha}d(x_i,Q_j) \qquad (6)$$

که $U=(u_{ij})_{n\times k}$ ماتریس افراز فازی است که

$0 \le u_{ij} \le 1$ و $\sum_{j=1}^{k} u_{ij} = 1$.

$Q_j = (q_{j1}, q_{j2}, \cdots, q_{jp}, \tilde{v}_{jp+1}, \cdots, \tilde{v}_{jm})'$ بردار نماینده یا نمونه اولیه خوشه jام است. α $(1 < \alpha < \infty)$ ضریب فازی‌ساز است. همچنین $d(x_i, Q_j)$ اندازه عدم تشابه بین شیء x_i با نمونه اولیه Q_j است که به صورت زیر تعریف می‌شود:

$$d(x_i, Q_j) = \sum_{l=1}^{p} (w_l(x_{il}^r - q_{jl}^r))^2 + \sum_{l=p+1}^{m} \phi(x_{il}^c, \tilde{v}_{jl}^c)^2 \quad (۷)$$

بنابراین، می‌توان تابع هدف را به صورت زیر بازنویسی کنیم:

$$E(U,Q) = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ij}^{\alpha} \left(\sum_{l=1}^{p} (w_l(x_{il}^r - q_{jl}^r))^2 + \sum_{l=p+1}^{m} \phi(x_{il}^c, \tilde{v}_{jl}^c)^2\right) \quad (۸)$$

w_l با استفاده از رابطه (۵) محاسبه می‌شود و

$$u_{ij} = \left(\sum_{z=1}^{k} \left(\frac{d(x_i,Q_j)}{d(x_i,Q_z)}\right)^{\frac{1}{\alpha-1}}\right)^{-1} \quad (۹)$$

q_{jl}^r مرکز ویژگی عددی است که به صورت زیر تعریف می‌شود:

$$q_{jl}^r = \frac{\sum_{i=1}^{n} u_{ij}^{\alpha} x_{il}^r}{\sum_{i=1}^{n} u_{ij}^{\alpha}} \quad (۱۰)$$

$\tilde{v}_{jl}^c$ مرکز فازی برای ویژگی طبقه‌ای است که به صورت زیر تعریف می‌شود:

$$\tilde{v}_{jl}^c = \frac{a_{jl}^1}{w_{jl}^1} + \frac{a_{jl}^2}{w_{jl}^2} + \cdots + \frac{a_{jl}^k}{w_{jl}^k} + \cdots + \frac{a_{jl}^t}{w_{jl}^t} \quad (۱۱)$$

و

$$\phi(x_{il}^c, \tilde{v}_{jl}^c) = \sum_{k=1}^{t} \tau(x_{il}^c, a_{jl}^k) \times \delta(x_{il}^c, a_{jl}^k), \quad (۱۲)$$

که $\delta(x_{il}^c, a_{jl}^k)$ با استفاده از رابطه (٤) محاسبه می‌شود

و

$$\tau(x_{il}^c, a_{jl}^k) = \begin{cases} 0, & x_{il}^c = a_{jl}^k \\ \omega_{jl}^k, & x_{il}^c \neq a_{jl}^k \end{cases}, \quad (۱۳)$$

و

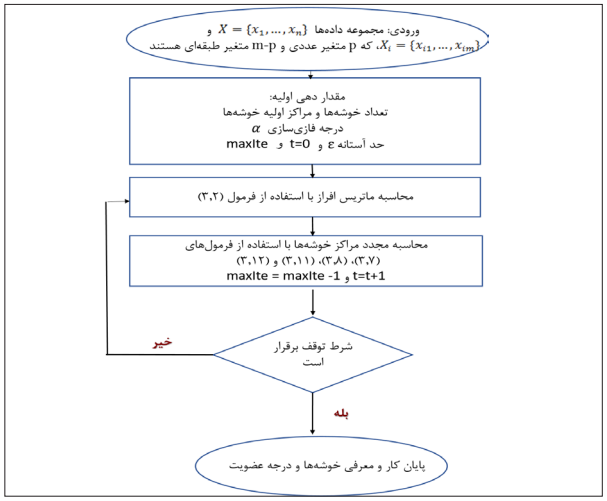
$$\tau(x_{il}^c, a_{jl}^k) = \begin{cases} 0, & x_{il}^c = a_{jl}^k \\ \omega_{jl}^k, & x_{il}^c \neq a_{jl}^k \end{cases},$$

$$\omega_{jl}^k = \sum_{i=1}^{n} \gamma(x_{il}^c) \quad (۱٤)$$

که

$$\gamma(x_{il}^c) = \begin{cases} \frac{u_{ij}^{\alpha}}{\sum_{i=1}^{n} u_{ij}^{\alpha}}, & x_{il}^c = a_{jl}^k \\ 0, & x_{il}^c \neq a_{jl}^k \end{cases}, \quad (۱۵)$$

به طورکلی می‌توان الگوریتم خوشه‌بندی فازی K-prototypes را به صورت شکل ۲ خلاصه کرد. پارامترهای این الگوریتم نیز در جدول ۱ آمده است.



شکل ۲: الگوریتم خوشه‌بندیK-prototypes فازی

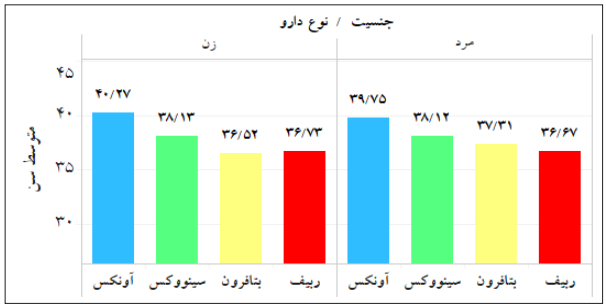
جدول ۱: پارامترهای الگوریتم فازی

پارامتر	مقدار
شاخص فازی ساز	۱/۵
مقدار تابع هدف	۷۸۰/۷۴
مقدار خطا	۱E-۱۵
تعداد تکرار	۲۸
$MaxIte$	۵۰۰

۵- معیارهای ارزیابی خوشه‌بندی

هدف خوشه‌بندی، یافتن گروه‌های مشابه از بین مجموعه داده ورودی است. یکی از مهمترین مراحل خوشه‌بندی، ارزیابی نتایج خوشه‌بندی است که از طریق آن، میزان مناسب بودن نتیجه خوشه‌بندی را می‌توان

بررسی کـــرد. در واقع، هیچ معیاری بـــه طور مطلق برای انتخاب بهترین خوشـــه‌بندی وجود نـــدارد. این انتخاب به ماهیـــت داده‌ها و دیدگاه کاربر بســـتگی دارد. حتی پس از انتخاب بهتریـــن روش، همچنان نیاز به قضاوت انســـانی وجود دارد. در این راستا، این پژوهش به معرفی سه روش ارزیابی رایج پرداخته اســـت کـــه در ادامه به طور خلاصه توضیح داده خواهد شـــد. این روش‌ها نه تنها بهینه‌سازی عملکرد خوشـــه‌بندی را تضمیـــن می‌کنند، بلکه همچنین به متخصصـــان داده در ارزیابی دقـــت و کارایی روش‌های مختلف کمک می‌کنند.



شکل ۳: متوسط سن بیماران به تفکیک نوع داروی مصرفی و جنسیت

شـــاخص ضریب نیم‌رخ[4] که به آن ضریب سیلوئت نیز گفته می‌شـــود، یک معیار پرکاربرد بـــرای ارزیابی کیفیت نتایج خوشـــه‌بندی اســـت. مقدار این شـــاخص بین ۱- تا ۱+ تغییـــر می‌کند. مقدار نزدیک به ۱، بیانگر این اســـت که یک نقطه به خوبی با خوشـــه خودش مطابقت داشـــته و با خوشـــه‌های مجاور همخوانی ضعیفی دارد. کوچک بودن ضریب سیلوئت برای خوشه‌ها، بیانگر ضعیف بودن نتایج خوشـــه‌بندی است که ممکن است به علت انتخاب نامناسب تعداد خوشه‌ها (k) نیز باشد [۲۲]. تفسیر مقادیر مختلف این شاخص در جدول ۲ ثبت شده است.

۲- شـــاخص دیویس- بولدین[5]، این شـــاخص میانگین شـــباهت بین هر خوشـــه را با شـــبیه‌ترین خوشـــه به آن محاســـبه می‌کند. هر چه مقدار شـــاخص دیویس- بولدین کمتر باشد، نتیجه خوشه‌بندی بهتر است، زیرا نشان می‌دهد

4-Silhouette coefficient
5-Davis–Bouldin

جدول۲: تفسیر مقادیر مختلف ضریب سیلوئت

تفسیر	مقدار ضریب سیلوئت
ساختار قوی	۷۱-۱۰۰ درصد
ساختار مناسب	۵۱-۷۰ درصد
ساختار ضعیف	۲۵-۵۰ درصد
هیچ ساختار قابل توجهی وجود ندارد	کمتر از ۲۵ درصد

که خوشه‌ها به خوبی از هم جدا شده و فشرده هستند. این شـــاخص به ویژه برای تعیین تعداد بهینه خوشه‌ها در یک مجموعه داده مفید است [۲۳].

۳- شاخص کالینســـکی- هاراباسز[6] نیز معیار دیگری برای بررسی کیفیت خوشه‌بندی است که مقدار بالاتر این شاخص نشـــان‌دهنده خوشه‌بندی بهتر است. این شاخص معمـــولاً در تحلیل خوشـــه‌ای بـــرای تعیین تعـــداد بهینه خوشـــه‌ها و ارزیابی اثربخشی الگوریتم‌های خوشه‌بندی استفاده می‌شود.

۶- مواد و روش کار

داده‌هـــای مورد اســـتفاده در این مطالعـــه‌ی مقطعی- تحلیـــلی از پایگاه داده پژوهش برجـــوئی و همکاران [۱۹]، با همکاری مستقیم پژوهشگران اصلی، در اختیار نویسنده قرار گرفـــت. پس از انجام فرایندهـــای دقیق پیش‌پردازش و پاک‌سازی شـــامل حذف موارد ناقص یا تکراری، تعداد ٤٤٨ رکورد معتبر انتخاب شـــد. این داده‌ها شامل اطلاعات جمعیت‌شناختی (محل تولد، سن، سن در زمان تشخیص) و دارویی بیماران مبتلا به مولتیپل اسکلروزیس در استان چهارمحال و بختیاری بوده و برای تحلیل‌های خوشه‌بندی مورد استفاده قرار گرفتند.

داروهـــای مصرفی بیمـــاران عبارتنـــد از سینووکس[7] (اینترفرون بتا-۱ ای) دارویی که به صورت عضلانی تزریق می‌شود، این دارو معمولاً در زمانی که بیماری ام.اس تشدید می‌شود تجویز شده و موجب کاهش حملات ناشی از آن و متعاقباً کند شدن روند پیشرفت بیماری می‌شود. آوونکس[8]

6-Calinski-Harabasz
7-Cinnovex (Interferon beta-1a)
8-Avonex

(اینترفرون بتا-۱ ای) دارویی که به صورت عضلانی تزریق می‌شود و از پیشرفت بیماری ام.اس و تشدید علائم آن (مانند مشکلات تعادل، بی‌حسی یا ضعف) جلوگیری می‌کند. ربیف[9] (اینترفرون بتا-۱ ای) دارویی است که برای درمان نوع عودکننده بیماری ام.اس به صورت تزریق زیر جلدی استفاده می‌شود، این دارو دفعات عود و شدت علائم بیماری را کاهش می‌دهد. بتافرون (اینترفرون بتا-۱ بی)[10] دارویی که به‌صورت زیرجلدی تزریق می‌شود و برای کاهش تعداد حملات عودکننده بیماری ام.اس تجویز می‌شود.

پیش‌پردازش داده‌ها: برای ارتقاء کیفیت داده‌ها، فرایند پیش‌پردازش شامل مراحل زیر به دقت انجام شد: ابتدا، داده‌های ناقص به دلیل حجم اندک و احتمال ایجاد سوگیری ناشی از روش‌های تکمیل داده مانند میانگین‌گذاری، حذف گردیدند. سپس، ویژگی‌های عددی شامل سن بیمار و سن در زمان تشخیص با استفاده از استانداردسازی Z-score نرمال‌سازی شدند تا اثر تفاوت مقیاس‌ها در تحلیل‌ها کاهش یابد. در نهایت، با به‌کارگیری روش فاصله بین‌چارکی (IQR) برای شناسایی داده‌های پرت، هیچ نمونه پرت قابل‌توجهی مشاهده نشد.

انتخاب روش: در این مطالعه، الگوریتم K-prototypes به‌عنوان روش اصلی خوشه‌بندی انتخاب شد، چرا که قابلیت پردازش هم‌زمان داده‌های عددی و طبقه‌ای را داراست. روش‌های جایگزین نظیر K-means و K-modes به‌ترتیب تنها برای داده‌های عددی و طبقه‌ای مناسب‌اند و در مواجهه با داده‌های ترکیبی، با محدودیت مواجه می‌شوند؛ همچنین، روش DBSCAN با وجود مزایای خاص، به دلیل حساسیت بالا به پارامترهای چگالی و عدم پشتیبانی از متغیرهای طبقه‌ای، کنار گذاشته شد. افزون بر این، از نسخه فازی الگوریتم K-prototypes بهره گرفته شد. روش فازی به دلیل انعطاف‌پذیری در تخصیص عضویت، برای تحلیل داده‌های بیماران ام.اس ترجیح داده شد، چرا که بیماران ممکن است هم‌زمان ویژگی‌هایی از چند زیرگروه مختلف را دارا باشند.

پیاده‌سازی محاسباتی: کلیه تحلیل‌های آماری و یادگیری ماشین با بهره‌گیری از محیط‌های برنامه‌نویسی Python (نسخه ۳,۹) و R انجام پذیرفت. این انتخاب مبتنی بر قابلیت‌های گسترده کتابخانه‌های تخصصی این نرم‌افزارها در پیاده‌سازی الگوریتم‌های خوشه‌بندی بود.

۷- یافته‌ها

۷۹ درصد (۳۵٤ نفر) از بیماران این پژوهش زن و بقیه مرد بوده‌اند. متوسط سن بیماران تقریبا ۳۸/۱۲ سال بوده به طوری که متوسط سن بیماران در زمان تشخیص بیماری تقریبا ۳۰ سال (متوسط سن زنان ۳۰/۱۰ و متوسط سن مردان، ۲۹/۵۸) بوده است. در شکل ۳ متوسط سن بیماران به تفکیک نوع داروی مصرفی و جنسیت آمده است.

در اجرای الگوریتم‌های K-prototypes فازی و غیرفازی، تعداد خوشه‌های بهینه، با توجه به حداکثر مقدار برای ضریب سیلوئت، شاخص‌های دیویس- بولدین و کالینسکی- هاراباسز، سه خوشه تعیین شد. ارزیابی عملکرد الگوریتم مزبور در جدول ۳ ارائه شده است. طبق این جدول، ضریب سیلوئت و شاخص کالینسکی- هاراباسز برای تعداد سه خوشه، مقادیر بیشتری را در مقایسه با دو خوشه به خود اختصاص داده‌اند. همین طور شاخص دیویس- بولدین برای سه خوشه دارای مقدار کمتری در مقایسه با دو خوشه است. بنابراین می‌توان گفت عملکرد دو روش خوشه‌بندی با در نظر گرفتن سه خوشه، تقریبا یکسان است.

جدول۳: معیارهای ارزیابی روش‌های خوشه‌بندی به ازای $k = 2$ و $k = 3$

روش خوشه‌بندی	k	ضریب سیلوئت	دیویس-بولدین	کالینسکی-هاراباسز
غیرفازی	۲	۰/۶۲	۰/۶۹	۴۲۰/۳۶
	۳	۰/۶۸	۰/۵۱	۷۹۴/۳۵
فازی	۲	۰/۶۱	۰/۸۴	۴۲۰/۹۷
	۳	۰/۶۶	۰/۶۷	۷۵۳/۶۱

9-Rebif
10-Rebif (Interferon beta-1b)

با توجه به تعداد زیاد نمونه‌ها، برای تعدادی از داده‌های نمونه درجه عضویت تعلق به هر خوشه با استفاده از روش خوشه‌بندی K-prototypes فازی در جدول ٤ ثبت شده است. ستون آخر این جدول، تعلق هر نمونه به یکی از سه خوشه را نشان می‌دهد.

جدول۴: درجه عضویت برخی از نمونه‌های موجود در داده‌ها برای هر خوشه

الگوریتم غیر فازی	شماره نمونه	خوشه اول	خوشه دوم	خوشه سوم
۱	۱	۰/۸۸۵	۰/۰۷۸	۰/۰۳۷۳
۲	۹۴	۰/۰۰۰۲	۰/۹۹۹۶	۰/۰۰۰۲
۳	۲۰۰	۰/۰۱۲	۰/۰۳۶	۰/۹۵۲
۲	۳۸۸	۰/۰۰۸	۰/۹۸۳	۰/۰۰۹
۱	۴۴۸	۰/۶۳۷	۰/۳۱۳	۰/۰۵

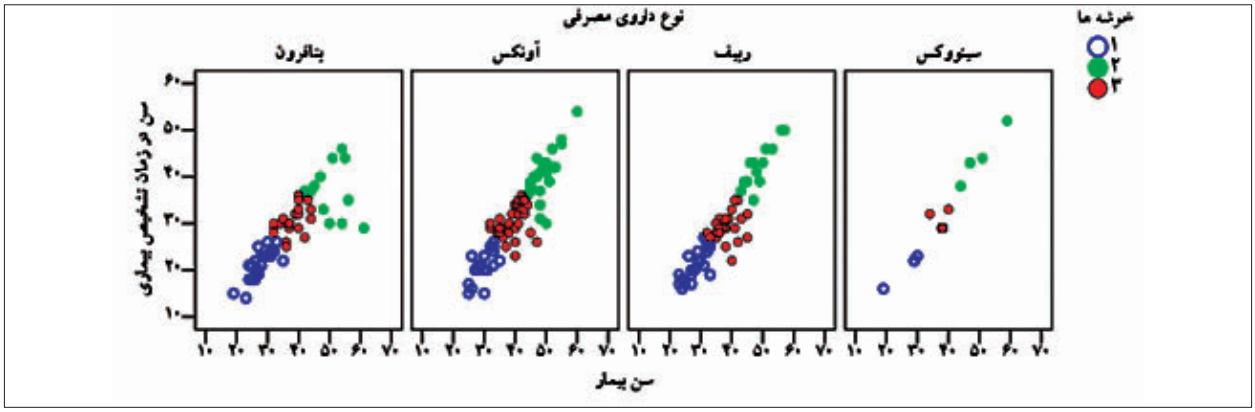
در روش خوشه‌بندی K-prototypes نیز خوشه‌های به دست آمده به صورت زیر توصیف می‌شوند:

خوشه یک: این خوشه شامل ١٦/٣ درصد رکوردهای مورد مطالعه بود، به طوری که ۷۸ درصد بیماران این خوشه را زنان با میانگین سنی ۲۸ سال و مابقی را مردان با میانگین سنی ۲۹ سال تشکیل می‌دادند. در این خوشه، کمترین سن بیمار ۱۹ و بیشترین سن ۳۵ سال بود. بیشتر افراد این خوشه، داروی ربیف و بتافرون (در مجموع حدود ۶۶ درصد) مصرف می‌کردند و درصد خیلی کمی داروی مصرفی شان سینووکس بوده است. متوسط سن بیماران در زمان تشخیص بیماری تقریبا ۲۱ سال بوده به طوری که کمترین سن در زمان تشخیص بیماری ۱۴ سال و بیشترین سن در زمان تشخیص بیماری ۲۷ سال بوده است.

خوشه دو: این خوشه شامل ۷۰/۱ درصد رکوردهای مورد مطالعه بود، به طوری که ۷۹ درصد بیماران این خوشه را زنان و مابقی را مردان تشکیل می‌دادند. سن بیماران این خوشه به طور متوسط ۳۸ سال بود به طوری که کمترین سن بیماران ۴۲ و بیشترین سن ۶۱ سال بود. بیشتر افراد این خوشه، داروی سینووکس (حدود ۷۱ درصد) مصرف می‌کردند و درصد خیلی کمی داروی مصرفی‌شان بتافرون بوده است. متوسط سن بیماران در زمان تشخیص بیماری تقریبا ۳۰ سال بوده به طوری که کمترین سن در زمان تشخیص بیماری ۲۲ سال و بیشترین سن در زمان تشخیص بیماری ٣٦ سال بوده است.

خوشه سه: این خوشه شامل ١٣/٦ درصد رکوردهای مورد مطالعه بود، به طوری که ۷۹ درصد بیماران این خوشه را زنان با میانگین سنی ٤٩ سال و مابقی را مردان با میانگین سنی تقریبا ٤٨ سال تشکیل می‌دادند. در این خوشه کمترین سن بیمار ٤٢ و بیشترین سن ٦١ سال بود. بیشتر افراد این خوشه داروی آونکس (حدود ٤٩ درصد) مصرف می‌کردند و درصد خیلی کمی داروی مصرفی شان سینووکس بوده است. متوسط سن بیماران در زمان تشخیص بیماری تقریبا ٤٠ سال بوده به طوری که کمترین سن در زمان تشخیص بیماری ۲۹ سال و بیشترین سن در زمان تشخیص بیماری ٥٤ سال بوده است.

شناسایی سه خوشه متمایز با ویژگی‌های دموگرافیک و دارویی خاص، نشان‌دهنده تنوع بالینی در بیماران ام.اس است و این تقسیم‌بندی می‌تواند به شناسایی الگوهای پنهان در میان بیماران کمک کرده و موجب تفکیک بهتر زیرنوع‌های بالینی بیماری شود [٢٤]. به‌عنوان نمونه، خوشه اول شامل بیماران جوان‌تر (میانگین سنی حدود ۲۸ سال) با مصرف غالب ربیف و بتافرون، و خوشه سوم شامل بیماران مسن‌تر (میانگین سنی ٤٩ سال) با مصرف غالب آوونکس بود. این تفاوت‌ها بیانگر ارتباط میان سن، جنسیت و نوع داروی مصرفی با سیر بیماری بوده و می‌تواند بینش‌های جدیدی برای طراحی رویکردهای درمانی شخصی‌سازی شده فراهم کند [٢٥]. همچنین، تفاوت قابل‌توجه سن در زمان تشخیص و انتخاب دارو در خوشه‌ها، نقش آغاز زود هنگام درمان‌های تعدیل‌کننده بیماری را در بهبود پیامدهای بالینی و کیفیت زندگی بیماران تأیید می‌کند [٢٦]. شناخت این خوشه‌ها به پزشکان امکان می‌دهد درمان‌ها را متناسب با پروفایل ترکیبی هر



شکل ۴: مصورسازی خوشه‌ها برحسب متغیرهای عددی و طبقه‌ای در روش خوشه‌بندی K-prototypes غیر فازی.

خوشه به‌صورت اختصاصی‌تر اعمال کنند. این رویکرد می‌تواند نیاز به داروهای با اثربخشی بالاتر یا عوارض کمتر را شناسایی کرده و برنامه‌ریزی فردمحور را تسهیل کند. در نتیجه، خوشه‌بندی به‌عنوان ابزار طبقه‌بندی بالینی، پایه تصمیم‌گیری درمانی مبتنی بر داده و همسو با پزشکی شخصی در بیماری‌های مزمن است [۲۷].

یافته‌های حاصل از خوشه‌بندی الگوریتمی، با ارائه داده‌های کمّی و قابل استناد، امکان اولویت‌بندی درمان و پایش دقیق‌تر سیر بیماری را بر اساس پروفایل‌های بالینی تجمیعی فراهم می‌کنند و از این طریق به افزایش دقت درمان، کاهش هزینه‌ها و ارتقاء کیفیت زندگی بیماران کمک می‌کنند [۲۸]. شواهد نشان می‌دهد که اثربخشی درمان‌های ایمونومدولاتوری در ام.اس به‌شدت وابسته به سن است و پس از ۵۳ سال تقریباً به صفر می‌رسد. بنابراین، خوشه‌بندی می‌تواند ابزار مهمی برای تدوین برنامه‌های درمانی متناسب با سن و ویژگی‌های دموگرافیک بیمار باشد. این امر به پزشکان امکان می‌دهد درمان‌های با کارایی بالا را برای بیماران جوان‌تر (مانند خوشه اول با میانگین سنی ۲۸ سال) و رویکردهای متمرکز بر مدیریت علائم را برای بیماران مسن‌تر (مانند خوشه سوم با میانگین سنی ۴۹ سال) در نظر گیرند، رویکردی که هم با شواهد کاهش اثربخشی درمان در سنین بالاتر و هم با ضرورت آغاز زودتر درمان برای بهبود نتایج بالینی هم‌راستا است. [۲۵, ۲۶].

۸- بحث و نتیجه‌گیری

این مطالعه با بهره‌گیری از الگوریتم‌های K-prototypes در دو حالت فازی و غیرفازی، به تحلیل خوشه‌ای بیماران مبتلا به مولتیپل اسکلروزیس در استان چهارمحال و بختیاری پرداخت. تحلیل‌ها بر مبنای ویژگی‌های جمعیت‌شناختی و دارویی انجام شد. با استفاده از شاخص‌های معتبر ارزیابی از جمله سیلوئت، دیویس- بولدین و کالینسکی- هاراباسز، سه خوشه بهینه شناسایی گردید. مقایسه عملکرد دو الگوریتم نشان داد که هر دو روش دقت بالایی دارند، با این حال روش فازی به‌دلیل امکان تخصیص درجات عضویت، تحلیل دقیق‌تری از تعلق بیماران به خوشه‌ها ارائه می‌دهد. این انعطاف‌پذیری، مزیتی مهم در مواجهه با ناهمگنی بالینی بیماران ام.اس به‌شمار می‌رود.

یافته‌های این مطالعه با پژوهش‌های مشابه داخلی و خارجی اسحاقی و همکاران [۱۶] و برانکو و همکاران [۱۵] همسو بوده و بر اهمیت تحلیل خوشه‌ای در طراحی پروتکل‌های درمانی شخص‌محور تأکید دارد. شناسایی الگوهای متمایز در زیرگروه‌های بیماران می‌تواند به عنوان چارچوب تصمیم‌گیری بالینی برای بهینه‌سازی تجویز دارو و مدیریت سیر بیماری مورد استفاده قرار گیرد.

با وجود نتایج قابل توجه، محدودیت‌هایی همچون حجم نمونه نسبتاً اندک و تمرکز جغرافیایی مطالعه وجود دارد. لذا پیشنهاد می‌شود در تحقیقات آینده از داده‌های

چندمرکزی و گســـترده‌تر بهره‌گیری شود و الگوریتم‌های پیشرفته‌تر یادگیری ماشین، همراه با ادغام داده‌های بالینی، تصویربـــرداری و ژنتیکی، به‌کار گرفته شـــود تا الگوهای پیچیده‌تر بیماری با دقت بیشتری شناسایی گردد.

مراجع

1. Reingold, S. C. 2002.Theoretical Concepts, Current Research, and Clinical Trials in Multiple Sclerosis. Comprehensive Nursing Care in Multiple Sclerosis 2002; 15.
2. Horn, D. 2001. Clustering via Hilbert space. Physica A: Statistical Mechanics and its Applications 2001; 302: 70–79.
3. Jain, A. K. 2009. Data clustering: 50 years beyond k-means. Pattern Recognition Letters 2009; 31(8): 651–666.
4. Zadeh, L. A. 1965. Fuzzy sets. Information and Control 1965; 8: 338–353.
5. Atanassov, KT. 1986. Intuitionistic fuzzy sets. Fuzzy Sets and Systems 1986; 20: 87–96.
6. Torra, V. 2010. Hesitant fuzzy sets. International Journal of Intelligent Systems 2010; 25: 529–539.
7. Ruspini, ER. 1969. A new approach to clustering. Information and Control 1969; 15(1): 22–32.
8. Dunn, J.C. 1973. A fuzzy relative of the isodata process and its uses in detecting compact well-separated clusters. Journal of Cybernetics 1973; 32–57.
9. Bezdek, J.C. & Dunn, J.C. 1986. Optimal fuzzy partitions; a heuristic for estimating the parameters in a mixture of normal distributions. IEEE Transactions on Computers 1986; C-35(8): 935–938.
10. فرنام، م. و دره‌میرکی، م. ۱۴۰۱. خوشه‌بندی داده‌های فازی با به کارگیری الگوریتم FCM بر مبنای یک معیار اندازه فاصله پارامتری. سیســـتم‌های فازی و کاربردها. ۵ (۲): ۱۱۹–۹۳.
11. Zhi, XB., Fan, JL. & Zhao, F. 2014. Robust local feature weighting hard c-means clustering algorithm. Neurocomputing 2014; 134: 20–29.
12. Wang, X., Wang, Y., & Wang, L. 2004. Improving fuzzy c-means clustering based on feature-weight learning. Pattern Recognition Letters 2004; 25: 1123–1132.
13. Xing, HJ., & Ha, MH. 2014. Further improvements in Feature-Weighted Fuzzy C-Means. Information Sciences 2014; 267: 1–15.
14. Wu, KL., & Yang, MS. 2002. Alternative c-means clustering algorithms. Pattern Recognition 2002; 35: 2267–2278.
15. Branco, D., Martino, BD., Esposito, A., Tedeschi, G., Bonavita, S., & Lavorgna, L. 2022. Machine learning techniques for prediction of multiple sclerosis progression. Soft Computing 2022; 26(22): 12041–12055.
16. Eshaghi, A., Young, AL., Wijeratne, PA., et al. 2021. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. Nature Communications 2021; 12: 2078.
17. Huang, Z. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery 1998; 2(3): 283–304.
18. Ji, J., Pang, W., Zhou, C., Han, X., & Wang, Z. 2012. A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. Knowledge-Based Systems 2012; 30: 129–135.
19. Barjui, SP., Reiisi, S., & Bayati, A. 2017. Human glutathione s-transferase enzyme gene variations and risk of multiple sclerosis in Iranian population cohort. Multiple Sclerosis and Related Disorders 2017; 17: 41–46.
20. Bezdek, JC.1981. Pattern Recognition with Fuzzy Objective Function Algorithms. 1st Edition, Plenum Press: New York.
21. Gustafson, DE. 1979. Kessel W. Fuzzy clustering with a fuzzy covariance matrix. Proceedings of IEEE Conference on Decision and Control, San Diego 1979; 761–766.
22. Bai, L., Liang, JY., & Dang, CY. 2011. An initialization method to simultaneously find initial cluster centers and the number of clusters for clustering categorical data. Knowledge-Based Systems 2011; 24(6): 785–795.
23. Wijaya, YA., & Kurniady, DA. 2021. Setyanto E, Tarihoran WS, Rusmana D, Rahim R. Davies Bouldin index algorithm for optimizing clustering case studies mapping school facilities. TEM Journal 2021; 10(3): 1099–1103.
24. Ontaneda, D., Thompson, AJ., Fox, RJ., & Cohen, JA. 2017. Progressive multiple sclerosis: prospects for disease therapy, repair, and restoration of function. The Lancet 2017; 389: 1357-1366.
25. Weideman, AM., Tapia-Maltos, MA., Johnson, K., Greenwood, M., & Bielekova, B. 2017. Meta-analysis of the age-dependent efficacy of multiple sclerosis treatments. Frontiers in Neurology 2017; 8: 577.
26. He, A., Spelman, T., & Manouchehrinia, A. 2023. Ciccarelli O, Hillert J, McKay K. Association between early treatment of multiple sclerosis and patient-reported outcomes: a nationwide observational cohort study. Journal of Neurology, Neurosurgery & Psychiatry 2023; 94: 284-289.
27. Lublin, FD., Reingold, SC., Cohen, JA., Cutter, GR., Sørensen, PS., Thompson, AJ., ... Polman, CH. 2014. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014; 83: 278-286.
28. Hult, KJ. 2017. Measuring the potential health impact of personalized medicine: evidence from MS treatments. National Bureau of Economic Research 2017; No. w23900