تجزیه تحلیل علایم بالینی بیماران مبتلا به اِسکلروزِ چندگانه با استفاده از داده کاوی

زهرا رئیسی[1*]
پانتهآ رمضان‌نژاد[2]
مرضیه احمدزاده[1]
شهرام ترحمی[3]

۱- گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز، ایران.
۲- گروه پزشکی قانونی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
۳- گروه نورولوژی، دانشگاه علوم پزشکی جندی شاپور، اهواز، ایران.

[*] نویسنده مسئول: شیراز، بلوار مدرس، دانشگاه صنعتی شیراز، گروه مهندسی کامپیوتر و فناوری اطلاعات
تلفن: ۰۹۱۳۰۶۸۷۱۰۲
E-mail: z.reisi@sutech.ac.ir

چکیده

دریافت: ۱۳۹۵/۰۹/۰۸ ویرایش: ۱۳۹۶/۰۱/۲۳ پذیرش: ۱۳۹۶/۰۱/۳۰ آنلاین: ۱۳۹۶/۰۱/۳۱

زمینه و هدف: تکنیک‌های داده کاوی از معروف‌ترین روش‌ها به‌منظور استخراج دانش و داده‌های نهان و ارزشمند در حیطه پزشکی می‌باشند و از مهمترین دستاوردهای آن می‌توان به کمک‌رسانی در زمینه تشخیص نوع بیماری و یا انتخاب نوع درمان اشاره کرد. هدف از این مطالعه پرداختن به دو مسئله شناخت علایم بالینی تاثیرگذار بر روی بیماری اِسکلروزِ چندگانه (Multiple sclerosis, MS) و بررسی میزان تاثیرگذاری فاکتورهای مختلف در این بیماری و رابطه بین این فاکتورها با انواع دسته‌بندی موجود بیماری بود.

روش بررسی: داده‌ها مربوط به بیماران اِسکلروزِ چندگانه مربوط به استان چهارمحال و بختیاری بود. این مطالعه پژوهشی در آزمایشگاه دانشگاه صنعتی شیراز همراه با همکاری تیم پزشکی انجام شد.

یافته‌ها: در این مطالعه در بررسی مسئله اول مشخص شد که بیشترین علایم تاثیرگذار در این بیماری علایم بالینی- بینایی بود و در بررسی مسئله دوم نتایج حاصل‌شده نشان داد که نسبت ابتلا زنان به مردان چهار برابر می‌باشد. در واقع ۷۰٪ از افراد مبتلا، با تحصیلات بالاتر از دیپلم در دسته عود کننده- بهبود یابنده قرار گرفته‌اند و ۶۲/۵٪ از افراد در محدوده سنی ۲۰ تا ۴۰ سال هستند.

نتیجه‌گیری: وجود برخی از علایم MS در افراد، موقت و گذراست و در بسیاری از موارد توسط افراد نادیده گرفته می‌شود که در صورت آگاهی از چگونگی شیوع علایم بالینی، می‌تواند یک اخطار برای افراد پیش از شروع دوره بحرانی بیماری باشد که به‌نوبه خود می‌تواند موجب تشخیص سریع‌تر، درمان موثرتر و تا حدی جلوگیری از پیشرفت بیماری شود.

کلمات کلیدی: مولتیپل اسکلروزیس، داده کاوی، دسته‌بندی، درخت تصمیم‌گیری.

مقدمه

امروزه با رشد فناوری در سازمان‌ها، به‌طور دایم حجم انبوهی از داده‌ها در مدت زمان اندکی ذخیره می‌شود. روشن است که استفاده از این داده‌ها و تجزیه و تحلیل آن‌ها تصمیم‌گیری را مشکل می‌سازد.

داده کاوی برای استخراج دانش از داده‌های با حجم زیاد به‌کار می‌رود.[1] داده کاوی، فرایند کشف الگوهای جالب، با مفهوم و کاربردی که در حجم عظیمی از داده‌ها پنهان شده است، می‌باشد.[2] در واقع داده کاوی با شناخت درست از گذشته، آینده را پیش‌بینی می‌کند.[3] داده کاوی در زمینه‌های مختلفی به‌کار گرفته می‌شود و یکی از زمینه‌هایی که در آن کاربرد بسیاری دارد علم پزشکی است. در علم پزشکی کشف الگوها برای تشخیص بیماری‌های جدید بسیار موثر می‌باشد. بازیابی داده‌ها به‌وسیله کامپیوتر می‌تواند کیفیت تصمیم‌گیری را بهبود بخشد و از خطاهای انسانی جلوگیری کند. هر چند تصمیم‌هایی که توسط انسان گرفته می‌شوند بیشتر مطلوب هستند اما وقتی با حجم زیادی از داده‌ها روبه‌رو باشیم تصمیمات انسانی عملکرد پایین‌تری را ارایه می‌دهند. همچنین کارایی و دقت

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

تصمیمات زمانی که انسان‌ها با حجم و فشار زیاد کاری مواجه هستند، کاهش می‌یابد.[۴]

به‌طورکلی روش‌های دسته‌بندی در داده کاوی در حوزه پزشکی کاربرد بیشتری دارند.[۵] این درحالی‌است که درخت تصمیم‌گیری (Decision tree) در حالت کلی به‌طور معمول دقت قابل‌قبولی ارایه می‌دهد.[۶] کارهای انجام‌شده در زمینه پزشکی بیشتر از تکنیک‌های دسته‌بندی استفاده کرده‌اند. در این مقاله به‌صورت خلاصه تعدادی از کارهای انجام‌شده را شرح می‌دهیم.

Sepehri و همکاران، "ساخت مدلی برای تعیین نوع درمان سنگ حالب" به‌منظور طراحی یک سیستم تصمیم یار جهت انتخاب نوع درمان سنگ کلیه با توجه به ویژگی‌های سنگ و بیمار ارایه کردند. در این روش از درخت تصمیم‌گیری استفاده شده است و با دقت ۸۵٪ نتایج حاصل به‌دست آمدند.[۷] در پژوهش پیش‌بینی احتمال ابتلا افراد به دیابت بر اساس فعالیت‌های سبک زندگی در آمریکا مطالعاتی انجام گرفته است و در نهایت الگوهای پیش‌بینی استخراج گردیده است.[۸] در "راهکاری جهت کشف روابط بین عوامل خطرزای قلبی" از داده کاوی برای به‌دست آوردن روابط غیر مفید بین عوامل خطرزا در بیماری‌های قلب و عروق استفاده شده است. تکنیک استفاده‌شده در این مقاله دسته‌بندی به‌کمک درخت‌های تصمیم‌گیری است.[۹] "تشخیص سکته مغزی به‌کمک تکنیک‌های داده کاوی" به‌منظور تشخیص و مدیریت موثرتر بیماری سکته مغزی توسط Ghafari و همکاران معرفی شده است و از دو تکنیک داده کاوی درخت تصمیم و نزدیک‌ترین همسایه برای دسته‌بندی داده‌ها استفاده شده است و دقت حاصل‌شده به‌ترتیب برابر با ۹۵/۴۲٪ و ۹۴/۱۸٪ می‌باشد.[۱۰]

در زمینه سرطان‌های مختلف با استفاده از تکنیک داده کاوی پیش‌بینی‌های مختلفی انجام گرفته است. Lihua و همکاران از روش‌های داده کاوی با استفاده از Serum proteomic profiling به پیش‌بینی سرطان اقدام کردند. این پژوهش معتقد است به‌خاطر پیچیدگی ناشی از Proteomic profiling آنالیز داده‌های زیاد نیاز به تکنیک‌های داده کاوی برای تشخیص الگوهای پیچیده دارد و در نهایت به تشخیص موفقیت‌آمیزی در سرطان تخمدان با کارایی قابل قبول دست یافتند.[۱۱]

کارهای مشابهی با استفاده از تکنیک‌های داده کاوی در پزشکی جهت کشف الگوهای موثر انجام شده است که از جمله آن‌ها می‌توان به سرطان پستان،[۱۲-۱۸] بررسی فاکتورهای موثر در ژن‌ها،[۱۹] بیماری‌های قلبی و عروقی،[۲۰-۲۶] سرطان پروستات[۲۷-۳۰] و دیابت[۳۱-۳۴] اشاره نمود.

لازم به یادآوری است در بیماری پارکینسون تکنیک‌های داده کاوی برای تشخیص نوسانات در بیماری ارایه گردیده است. در این مقاله بر روی تکنیک‌هایی تمرکز شده است که بر روی علایم بالینی اعمال می‌شوند و در نهایت الگوهای قابل قبولی ارایه می‌گردد.[۳۵]

در رابطه با بیماری‌های مرتبط با مغز کارهای ارزشمندی در زمینه داده کاوی صورت پذیرفته است. Walker و همکاران استراتژی‌هایی جهت برخورد با مشکلات آلزایمر ارایه کرده و دو کلاس بیماران مبتلا به آلزایمر و نرمال را تعریف کردند و به نتایج مفیدی دست یافتند که در پیش‌بینی آلزایمر و کمک به بیماران مبتلا تاثیرگذار بوده است.[۳۶]

بیماری اِسکلروزِ چندگانه جزو بیماری‌های Demye linative می‌باشد که به آن دسته از بیماری‌هایی گفته می‌شود که در زمان جنینی میلین به‌طور کامل و طبیعی ساخته شده ولی در طول زندگی توسط عواملی بیماری‌زا آسیب دیده‌اند. در ضمن یادآوری می‌شود که خود میلین (Myelin) ماده‌ای است از جنس فسفولیپید که آکسون نرون‌ها را در سیستم اعصاب مرکزی و محیطی می‌پوشاند.[۳۷]

در واقع اِسکلروزِ چندگانه بیماری مزمنی است که با درگیری نواحی مختلف سیستم عصبی مرکزی موجب بروز علایم بالینی متفاوت و ناتوانی بیمار می‌گردد. مکانیسم اصلی بیماری، به‌احتمالی بروز پدیده خودایمنی و تخریب میلین و تا حد کمتری آکسون‌های سیستم عصبی مرکزی است. واکنش سیستم ایمنی هم به‌صورت سلولی و هم تولید آنتی‌بادی‌های مختلف صورت می‌گیرد. وجه مشخص بیماری اِسکلروزِ چندگانه وجود پلاک‌های متعدد در بافت سفید مغز، ساقه مغز و نخاع می‌باشد که در این پلاک‌ها میلین به‌صورت کامل یا ناقص تخریب شده است. درحال‌حاضر با وجود پیشرفت‌های خیره‌کننده پزشکی درمان قطعی برای این بیماری یافت نشده است. اما طی چند سال اخیر استفاده از داروهای تاثیرگذار بر سیستم ایمنی، موجب کاهش حملات و تا حدی میزان ناتوانی بیماران شده است.[۳۸]

از نظر سیر بالینی، بیماری اِسکلروزِ چندگانه به چهار نوع مختلف تقسیم‌بندی شده است که عبارتند از:[۳۹]

فرم عود کننده–بهبودیابنده Relapsing-remitting MS, RRMS: در آن بیماران دچار حملاتی می‌شوند که به‌طور معمول چند روز تا چند هفته طول کشیده و سپس علایم به‌صورت کامل یا ناکامل بهبود

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

می‌یابند. در شروع بیشتر بیماران اِسکلروزِ چندگانه در این گروه قرار دارند (۸۵-۸۰٪).

پیشرونده ثانویه Secondary progressive MS, SPMS: بعضی از بیماران دسته اول (RRMS) با گذشت زمان تبدیل به این نوع بیماری می‌شوند. در این حالت، بیماری بدون وجود حمله آشکار به آرامی شروع به پیشرفت می‌نماید. برخی از این بیماران با وجود پیشرفت علایم طی زمان حملاتی را نیز تجربه می‌کنند و در سیر پیشرفت، گاهی بهبودهای خفیف دارند یا زمان‌هایی را در حالت ثبات به‌سر می‌برند.

پیشرونده اولیه Primary progressive MS, PPMS: در این فرم علایم بیمار از ابتدا پیشرونده بوده و بیمار حمله‌ای را تجربه نمی‌کند.

پیشرونده–عودکننده Progressive-relapsing MS, PRMS: بیماری از ابتدا حالت پیشرونده همراه با عودهای حاد دارد که می‌تواند بدون بهبود کامل حملات باشد. در ضمن بین حملات، بیماری سیر پیشرونده دارد.

گفتنی است که بیمارانی که تنها یک حمله داشته باشند تحت عنوان سندرم بالینی مجزا یا Clinically isolated syndrome نامیده می‌شوند که ممکن است به سمت اِسکلروزِ چندگانه پیشرفت نمایند و حتی ممکن است بیمار سال‌ها در این مرحله بماند و علایم تکرار نشود. در این مطالعه بر این گونه بیماری در کنار چهار نوع بیماری اِسکلروزِ چندگانه نیز تمرکز شده است.

در رابطه با کارهای انجام‌گرفته در بیماری اِسکلروزِ چندگانه با تکنیک‌های داده کاوی، Li و همکاران مطالعه‌ای به‌منظور بررسی کیفیت زندگی در بیماران مبتلا به مولتیپل اسکلروزیس در آمریکا ارایه دادند. در این پژوهش تکنیک‌های تجزیه و تحلیل، درخت پارتیشن‌بندی بازگشتی انجام شده است و داده‌ها از طریق پرسشنامه استاندارد SF-12v2 و به روش مصاحبه جمع‌آوری گردیده است. این پرسشنامه دامنه سلامت جسمانی و سلامت روان را شامل می‌شود. هر دامنه امتیازاتی دارد که در برگ‌های درخت نشان داده می‌شوند. این پژوهش شامل ۳۰۰۸ رکورد داده بیمار با متوسط سن ۴۲/۵ می‌باشد. لازم به یادآوری است که ۸۳٪ بیماران زن می‌باشند. همچنین در این مطالعه نشان داده شده که تشخیص در کمتر از یک سال میزان پیشرفت را کند می‌کند.[۴۰]

Gironi و همکاران سیستم مبتنی بر داده کاوی جهت تشخیص رابطه بین عوامل ایمنی در بیماری MS ارایه دادند. در این کار از شبکه عصبی مصنوعی برای یافتن ارتباط بین متغیرها استفاده شده است.[۴۱] هدف این پژوهش کمک به پزشکان جهت تشخیص سریع بیماری و تجویز مناسب، و همچنین دادن نوعی پیش‌آگهی به افراد جهت جلوگیری از ابتلا به این بیماری بود.

روش بررسی

هدف این مطالعه یافته‌های بالینی به‌منظور کمک به تصمیم نهایی پزشک بود چراکه بر اساس معیارهای مک دونالد (Mac Donald) وجود حداقل یک علامت بالینی برای تشخیص MS ضروری است.[۴۲]

در پایگاه داده مورد استفاده این مقاله تمامی آزمایشات پاراکلینیک انجام شد و با فرض داشتن نتیجه تصویرسازی تشدید مغناطیسی (Magnetic resonance imaging, MRI)، بر روی یافته‌های بالینی بررسی انجام شد.

اولین مسئله‌ای که در این مطالعه به آن پرداخته شد شناخت و بررسی علایم بالینی تاثیرگذار در بیماری اِسکلروزِ چندگانه بود. سپس با توجه به شیوع این علایم، میزان تاثیرگذاری آن‌ها بر روی قطعیت یا عدم قطعیت تشخیص بیماری، بررسی شد. هدف از طرح این مسئله شناسایی علایم بالینی موثر در این بیماری جهت کمک به تشخیص سریع‌تر بیماری و تا حدودی جلوگیری از پیشرفت این علایم و در نهایت پیشرفت بیماری اِسکلروزِ چندگانه بود.

دومین مسئله که در این مطالعه بر روی آن کار تمرکز شده است بررسی میزان تاثیرگذاری فاکتورهای سن، جنس و میزان تحصیلات در بیماری اِسکلروزِ چندگانه بود. در این بیماری عوامل مختلفی مانند عوامل محیطی، ژنتیکی، اجتماعی و غیره می‌توانند نقش داشته باشند. در نهایت ضمن بررسی تاثیرگذاری سن و جنس و میزان تحصیلات، روابط بین این فاکتورها با چهار دسته‌بندی بیماری شناسایی می‌گردد.

داده‌های جمع‌آوری‌شده مربوط به بیماران استان چهارمحال و بختیاری می‌باشند که در دو بخش علایم بالینی و نتایج آزمایشات پاراکلینیک سازماندهی شده‌اند. همه بیماران دارای فیلدهای تشخیص، نوع بیماری، سن، جنسیت و میزان تحصیلات می‌باشند. همچنین برخی بیماران ثبت‌شده در این پایگاه داده دارای فیلد تشخیص قطعی بیماری می‌باشند.

علایم بالینی شامل موارد زیر است:

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

یافته‌های بالینی-بینایی: نوریت اپتیک، دوبینی، نیستاگموس، دیسمتری چشمی، افتالموپلژی بین‌هسته‌ای، جرقه/مگس‌پران، اختلال رفلکس مردمک.

یافته‌های بالینی-حرکتی: مونوپارزی، پاراپارزی، همی‌پارزی، کوادری پارزی، پاراپلژی، همی‌پلژی، کوادری پلژی، اسپاسم، دیزارتری، آتروفی عضلانی، کرامپ، هایپوتونی، میوکمی، سندرم پای بی‌قرار، رفلکس بابنسکی، رفلکس هوفمن، افتادگی پا.

یافته‌های بالینی-حسی: پارستزی، بی‌حسی، نورالژی، علامت لرمیت، اختلالات پروپریوسپتیو.

یافته‌های بالینی-رفتاری/ روانی: اختلالات خواب، تکرار ادرار، اختلالات نعوظ در مردان، انزال رتروگرید، سرد مزاجی، یبوست، اختلالات اجابت مزاج، ریفلاکس معده به مری، خستگی.

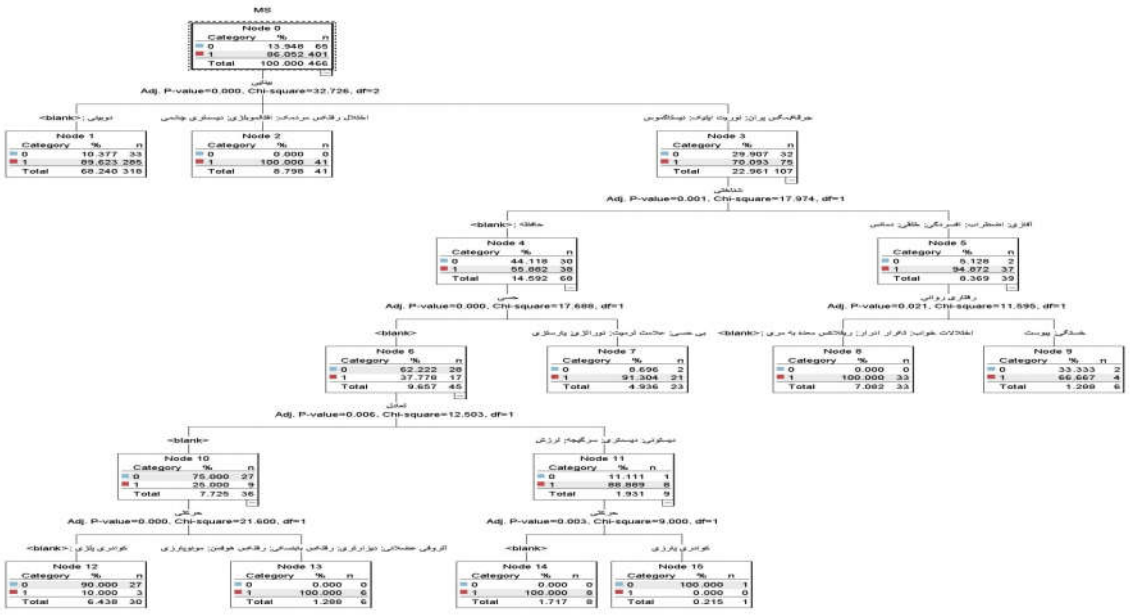
وجود فیلد خالی در علایم بالینی به‌منزله این است که بیمار این علامت را ندارد. در هر دو مسئله ۸۰٪ از داده‌ها به‌عنوان داده آموزشی و ۲۰٪ دیگر را به‌عنوان داده آزمایشی در نظر گرفته شده است. با توجه به کارایی بالای استراتژی‌های دسته‌بندی و دقت بالای نتایج، در این مطالعه برای دسته‌بندی از درخت تصمیم‌گیری استفاده گردیده است. در ابتدا فرآیند پیش‌پردازش جهت آماده‌سازی داده‌ها برای ساخت مدل انجام گرفته است. پس از آن درخت تصمیم‌گیری مربوط به نمونه‌ها ایجاد گردیده است و در نهایت کارایی مدل ساخته شده مورد ارزیابی قرار می‌گیرد. برای انجام تکنیک‌های داده کاوی از نرم‌افزار SPSS-Clementine software (http://www.the-data-mine.com/Software/SPSSClementine) استفاده شده است.

یافته‌ها

برای بررسی این مسئله از تکنیک دسته‌بندی و درخت تصمیم‌گیری استفاده شده است. الگوریتم مورد استفاده Chi-squared automatic interaction detection (CHAID) می‌باشد. فیلد هدف این مسئله قطعیت یا عدم قطعیت تشخیص بیماری اسکلروز چندگانه قرار داده شد.

در شکل ۱ نتیجه کلی از بررسی علایم بالینی نشان داده شده است. مهمترین متغیر در رسیدن به هدف، علایم بینایی هستند. پس از آن علایم تعادلی در این نتایج موثر بوده است و اثر علایم شناختی و حرکتی به‌طور تقریبی به یک اندازه است و علایم رفتاری-روانی نقش کمرنگ‌تری ایفا کرده‌اند.

با توجه شکل ۱ به‌طور مختصر نحوه تصمیم‌گیری و جهت‌دهی



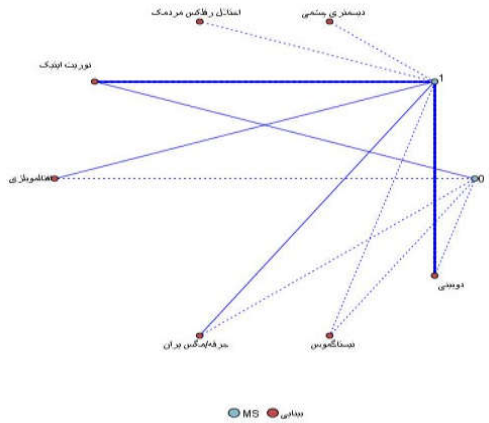
شکل ۱: بررسی علایم بالینی

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

علایم بالینی شرح داده می‌شود: درخت تصمیم‌گیری ابتدا با توجه به درصد و شیوع علایم بالینی اولین سطح دسته‌بندی را انجام داده و سه نوع از علایم که بیشترین شیوع را داشته‌اند و مربوط به علایم بالینی-بینایی هستند را برای ادامه تصمیم‌گیری انتخاب می‌کند. در سطح دوم با شیوع بالای علایم بالینی-شناختی مواجه می‌شود که بر اساس میزان شیوع علایم مربوط به این دسته، دو نوع تقسیم‌بندی را برای ادامه کار انتخاب می‌کند. این درخت در سطح سوم با دو دسته از علایم بالینی-حسی و علایم بالینی-رفتاری روانی مواجه می‌گردد و پس از بررسی، علایم بالینی رفتاری-روانی از ادامه کار برای تصمیم‌گیری حذف و علایم بالینی-حسی ادامه داده می‌شوند. سطح سوم درخت به تصمیم‌گیری در مورد علایم بالینی-تعادل و سطح چهارم به بررسی علایم بالینی-حرکتی می‌پردازد.

شکل ۲ میزان شیوع علایم و تاثیرگذاری آن‌ها را نشان می‌دهد. خطوط پر رنگ در این شکل نشان‌دهنده شیوع بیشتر علایم است و با کم شدن ضخامت خط، تاثیرگذاری علایم هم کمتر می‌شود. علامت‌های نوریت اپتیک و دوبینی بیشترین شیوع را نشان داده‌اند و سایر علایم قرار گرفته در این دسته به‌ترتیب اهمیت در دو سطح قرار می‌گیرند که در سطح اول علایم افتالموپلژی و جرقه/ مگس‌پران و در سطح دوم علایم دیسمتری چشمی، اختلال رفلکس مردمک و نیستاگموس قرار می‌گیرند.

جدول ۱ شامل نتایـج نهایی مدل استخراج شده می‌باشد. از این جدول برای بررسی درستی این مدل استفاده شده است. همان‌طور که از جدول استنباط می‌شود، الگوریتم CHAID، ٤٦١ رکورد از داده‌ها را برای قسمت آموزش درخت تصمیم‌گیری و ۱۲۲ رکورد برای قسمت تست در نظر گرفته است. در قسمت آموزش الگوریتم در ٤٧ رکورد پیش‌بینی اشتباه داشته است و در قسمت تست در ۱۷ رکورد این خطا دیده می‌شود. در نهایت دقت مدل آموزشی نزدیک به ۹۰٪ و دقت مدل آزمایشی حدود ٨٦٪ است که در واقع دقت قابل قبولی محسوب می‌شود.



شکل ۲: شیوع علایم بالینی-بینایی

در معرفی مسئله اول عنوان شد که در این مسئله بررسی ترکیب علایم بالینی نیز مورد بررسی ارزیابی قرار گرفته است. در مورد ترکیب علایم بالینی، شش ترکیب مهم از این مدل استخراج گردیده است. این ترکیبات با توجه به شیوع و میزان تاثیرگذاری علایم بالینی به‌دست آمده‌اند. این دسته‌بندی به‌صورت زیر می‌باشد:

۱- شناختی و حرکتی، ۲- تعادل و حرکتی، ۳- تعادل و شناختی، ٤- بینایی و شناختی، ۵- بینایی و حرکتی، ٦- بینایی و تعادل.

در مدل استخراج شده برای هر کدام از این ترکیبات، جدول‌ها و نمودارها به‌صورت جداگانه ترسیم شده است و مورد تجزیه و تحلیل قرار گرفته‌اند و در جدول ۲ ترکیب این علایم به تفکیک ارایه شده است. لازم به یادآوری است که این علایم با توجه به پایگاه داده موجود استخراج شده‌اند.

برای بررسی این مسئله از الگوریتم Apriori استفاده شده است. فیلد هدف این مسئله نوع بیماری می‌باشد. نوع بیماری شامل چهار دسته گفته‌شده در بخش پیشین می‌باشد. در ابتدا تاثیر جنسیت بر روی نوع بیماری مورد بررسی قرار می‌گیرد. همان‌طور که در جدول ۳ مشاهده می‌شود درصد ابتلا زنان نسبت به مردان چهار برابر است و ۵/٦۵٪ از زنان مبتلا به بیماری و ۳/٦٦٪ از مردان مبتلا در دسته عودکننده-بهبودیابنده قرار می‌گیرند.

با توجه به جدول ٤ افرادی که به‌طور متوسط تحصیلات بین دیپلم و بالاتر از دیپلم داشتند بیشترین گروه ابتلا به اِسکلروزِ چندگانه

جدول ۱: بررسی صحت مدل

بخش	تست		آموزش	
صحیح	۱۰۵	٪۸٦/۰۷	٤١٤	٪۸۹/۰۸
غلط	۱۷	٪۱۳/۹۳	٤٧	٪۱۰/۲
تعداد کل	۱۲۲		٤٦١	

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

جدول ۲: شیوع علایم

ترکیب علایم	درصد شیوع	ترکیب علایم	درصد شیوع
دوبینی و آتاکسی	۳٤/٦٥	اضطراب و اسپاسم	۲٤/۷٥
دوبینی و سرگیجه	۲۹/۷۰	اضطراب و پاراپارزی	۱۹/۸۰
دوبینی و لرزش	۲٤/۷٥	افسردگی و اسپاسم	۳۹/٦۰
دوبینی و اسپاسم	۲٤/۷٥	افسردگی و رفلکس بابنسکی	۱۹/۸۰
دوبینی و مونوپارزی	۱۹/۸۰	افسردگی و همی پارزی	۲٤/۷٥
دوبینی و همی پارزی	۲٤/۷٥	افسردگی و پاراپارزی	۲۹/۷۰
دوبینی و کرامپ	۱۹/۸۰	افسردگی و کرامپ	۲٤/۷٥
دوبینی و اضطراب	۲۹/۷۰	آتاکسی و افسردگی	٥۹/٤۰
دوبینی و افسردگی	٥٤/٤٥	آتاکسی و خلقی	۱۹/۸۰
نوریت اپتیک و آتاکسی	۹۰	سرگیجه و پاراپارزی	۲٤/۷٥
نوریت اپتیک و سرگیجه	۸۹/٥	سرگیجه واسپاسم	۲٤/۷٥
نوریت اپتیک و اسپاسم	٤۹/٥۰	لرزش و حافظه	۱٤/۸٥
نوریت اپتیک و رفلکس بابنسکی	۲٤/۷٥	لرزش و افسردگی	۱۹/۸۰
نوریت اپتیک و پاراپارزی	۳۹/٦۰	لرزش و پاراپارزی	۲٤/۷٥
نوریت اپتیک و کوادری پلژی	۱۹/۸۰	لرزش و کوادری پلژی	۳٤/٥٦
نوریت اپتیک و اضطراب	٤۹/٥۰	سرگیجه و همی پارزی	۲٤/۷٥
نوریت اپتیک و افسردگی	۸۸/٤٥	آتاکسی و اسپاسم	۳٤/٦٥
نوریت اپتیک و خلقی	۱۹/۸۰	افتالموبلژی بین‌هسته‌ای و افسردگی	۱۹/۸۰

جدول ۳: تاثیر جنسیت بر بیماری اسکلروز چندگانه

جنسیت	غیرقطعی بودن تشخیص	سندرم بالینی جداشده	عودکننده-بهبودیابنده	پیشرونده اولیه	پیشرونده ثانویه	پیشرونده-عودکننده
زن	۷۰	۱۰	۲٦۲	٤٤	٤۲	٤۲
مرد	۱۸	٥	٦۳	۱۱	۸	۸

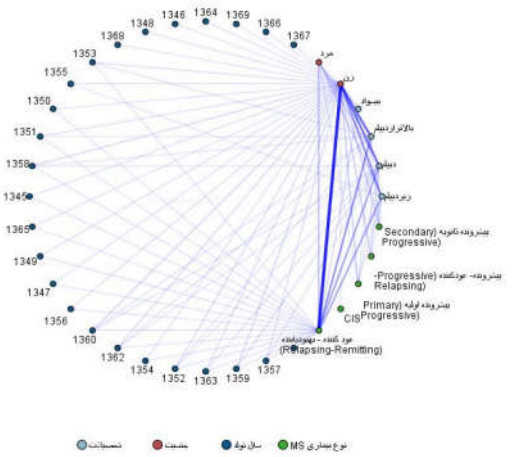
جدول ٤: تاثیر تحصیلات بر بیماری اسکلروز چندگانه

تحصیلات	غیرقطعی بودن تشخیص	سندرم بالینی جداشده	عودکننده-بهبودیابنده	پیشرونده اولیه	پیشرونده ثانویه	پیشرونده-عودکننده
بی‌سواد	۲	۰	۲٤	٤	٦	۷
زیردیپلم	۲۷	۲	۹۳	۱٤	۱٤	۱۰
دیپلم	۳۰	٥	۹٤	۲۳	۱۳	۲٤
بالاتر از دیپلم	۲۹	۸	۱۱٤	۱٤	۱۷	۹

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

جدول ۵: تاثیر سن بر بیماری اِسکلروزِ چندگانه

رده سنی	غیرقطعی بودن تشخیص	سندرم بالینی جداشده	عودکننده–بهبودیابنده	پیشرونده اولیه	پیشرونده ثانویه	پیشرونده–عودکننده
۵۰ تا ۶۰	۴	۰	۱۶	۰	۴	۲
۴۰ تا ۵۰	۱۷	۰	۷۹	۱۱	۶	۹
۳۰ تا ۴۰	۴۱	۷	۱۲۶	۱۷	۲۰	۲۴
۲۰ تا ۳۰	۲۶	۸	۱۰۴	۲۷	۲۰	۱۵



شکل ۳: ارتباط فاکتورهای مختلف با نوع بیماری

را تشکیل داده‌اند. در واقع ۷۰٪ از افراد با تحصیلات بالاتر از دیپلم در دسته عودکننده–بهبودیابنده قرار گرفته‌اند.

برای نشان دادن تاثیر سن بر روی این بیماری از جدول ۵ استفاده استفاده شده است. همان‌طور که از جدول استنباط می‌شود ۸۷/۸۸٪ از مبتلایان در رده سنی ۲۰ تا ۴۰ سال هستند و ۶۲/۵٪ از افراد مبتلای ۲۰ تا ۴۰ سال در دسته عودکننده–بهبودیابنده قرار گرفته‌اند.

در شکل ۳ گراف کلی ارتباط بین سن، جنس و میزان تحصیلات با نوع بیماری ارایه شده است. در این گراف نشان داده می‌شود که جنسیت با دسته عودکننده–بهبودیابنده بیشترین ارتباط را دارد و میزان تحصیلات بالاتر از دیپلم نیز به این دسته اضافه می‌شود. سال تولد افراد در این گراف با توجه به تاریخ خورشیدی می‌باشد و تاریخ تولد بین سال‌های ۱۳۵۱ تا ۱۳۷۱ بیشترین شیوع را نشان می‌دهند. در نتیجه افراد با ویژگی‌های زن بودن، سنین بین ۲۰ تا ۴۰ سال و میزان تحصیلات بالاتر از دیپلم بیشترین مبتلایان به بیماری اِسکلروزِ چندگانه را در این مطالعه تشکیل داده‌اند.

بحث

با توجه به نتایج به‌دست‌آمده از بررسی شناخت علایم بالینی تاثیرگذار بر روی بیماری اِسکلروزِ چندگانه در این پژوهش می‌توان پرخطرترین علایم این بیماری را در هر استان شناسایی کرد. همچنین با توجه به بررسی میزان تاثیرگذاری فاکتورهای سن، جنس و میزان تحصیلات در بیماری اِسکلروزِ چندگانه و رابطه بین این فاکتورها با انواع دسته‌بندی بیماری استنباط می‌شود که زنان در سنین بین ۲۰ تا ۴۰ سال و میزان تحصیلات بالاتر از دیپلم بیش از سایر افراد در معرض ابتلا به بیماری هستند.

نتیجه این مطالعه بیان‌کننده توجه و تمرکز بیشتر بر روی پیش‌آگهی این بیماری به‌خصوص قشر زنان در سنین جوانی می‌باشد. رابطه بین فاکتورهای موثر می‌تواند به تصمیم‌گیری‌های مدیریتی در این زمینه به مسئولین کمک نماید. همچنین می‌توان افراد مبتلا به بیماری را در دسته‌های مختلف بیماری مورد ارزیابی تخصصی قرار داد و راه‌حلی جهت جلوگیری از پیشگیری بیماری و انتقال به دسته‌های عودکننده یافت.

سپاسگزاری: این مقاله حاصل تلاش‌های معاونت دارویی استان چهارمحال و بختیاری جهت فراهم کردن داده‌های مورد نیاز پژوهش و همکاری دانشگاه صنعتی شیراز می‌باشد. نویسندگان مراتب تشکر و قدردانی خود را از این عزیزان به‌عمل می‌آورند.

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

References

1. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. *University of Illinois at Urbana-Champaign*; 2006.
2. Ngai EW, Xiu L, Chau DC. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst Appl* 2009;36(2):2592-2.
3. Wu X, Zhu X, Wu GQ, Ding W. Data mining with big data. *IEEE Trans Knowl Data Eng* 2014;26(1):97-107.
4. Eapen AG. Application of Data Mining in Medical Application. [Thesis]. Ontario, Canada: University of Waterloo, 2004.
5. Dobre OA, Abdi A, Bar-Ness Y, Su W. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Commun* 2007;1(2):137-56.
6. Anderson DR, Sweeney DJ, Williams TA, Camm JD, Cochran JJ, Ohlmann JW. An Introduction to Management Science: Quantitative Approaches to Decision Making. 14th ed. Mason, OH: South-Western Cengage Learning; 2014.
7. Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. A data mining based model for selecting type of treatment for kidney stone patients. *Tehran Univ Med J* 2009;67(6):421-7.
8. Repalli P. Prediction on Diabetes Using Data mining Approach. *Oklahoma State University,* 2011.
9. Kajabadi A, Saraee MH, Asgari S. Medical Data Mining: An Approach to Discovery Relationships Among Cardiovascular Risk Factors. 3rd Data Mining Conference Amir Kabir University, 2009.
10. Ghafari H, Amini L, Faraahi A. Stroke Diagnosis Using Data Mining Techniques. 5th Data Mining Conference Amir Kabir University, 2011.
11. Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, et al. Data mining techniques for cancer detection using serum proteomic profiling. *Artif Intell Med* 2004;32(2):71-83.
12. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005;34(2):113-27.
13. Bellaachia A, Guven E, Predicting breast cancer survivability using data mining techniques. *Age* 2006;58(13):10-110.
14. Grzymała-Busse J, Hu M. A Comparison of several approaches to missing attribute values in data mining. In: Ziarko W, Yao Y, editors. Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science. Berlin: Springer Heidelberg; 2001. P. 378-85.
15. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 2011;39(Database issue):D945-50.
16. Kuo WJ, Chang RF, Chen DR, Lee CC. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast Cancer Res Treat* 2001;66(1):51-7.
17. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* 2013;45(4):371-84, 384e1-2.
18. Chou SM, Lee TS, Shao YE, Chen IF. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Exp Syst Appl* 2004;27(1):133-42.
19. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 2004;6(1):1-6.
20. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques; Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA2008); 2008 Mar 31-Apr 4; Doha, Qatar. p. 108-15.
21. SA S, Intelligent heart disease prediction system using data mining techniques. *Int J Healthcare Biomed Res* 2013;1:94-101.
22. Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. *Int J Comput Appl* 2011;17(8):43-8.
23. Kumari M, Godara S, Comparative study of data mining classification methods in cardiovascular disease prediction, 2011; 1.
24. Dangare CS, Apte SS. Improved study of heart disease prediction system using data mining classification techniques. *Int J Comput Appl* 2012;47(10):44-8.
25. Soni J, Ansari U, Sharma D, Soni S. Intelligent and effective heart disease prediction system using weighted associative classifiers. *Int J Comput Sci Eng* 2011;3(6):2385-92.
26. Patil SB, Kumaraswamy YS. Extraction of significant patterns from heart disease warehouses for heart attack prediction. *IJCSNS* 2009;9(2):228-235.
27. Zhang H, Dong Y, Zhao H, Brooks JD, Hawthorn L, Nowak N, et al. Microarray data mining for potential selenium targets in chemoprevention of prostate cancer. *Cancer Genomics Proteomics* 2005;2(2):97-114.
28. Shah S, Kusiak A. Cancer gene search with data-mining and genetic algorithms. *Comput Biol Med* 2007;37(2):251-61.
29. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* 2007;39(1):41-51.
30. Loberg RD, Day LL, Harwood J, Ying C, St John LN, Giles R, et al. CCL2 is a potent regulator of prostate cancer cell migration and proliferation. *Neoplasia* 2006;8(7):578-86.
31. Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. *Artif Intell Med* 2002;26(1):37-54.
32. Sigurdardottir AK, Jonsdottir H, Benediktsson R. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. *Patient Educ Couns* 2007;67(1-2):21-31.
33. Koh HC, Tan G. Data mining applications in healthcare. *J Healthc Inf Manag* 2005;19(2):64-72.
34. Georga E, Protopappas V, Guillen A, Fico G, Ardigo D, Arredondo MT, et al. Data mining for blood glucose prediction and knowledge discovery in diabetic patients: the METABO diabetes modeling and management system. *Conf Proc IEEE Eng Med Biol Soc* 2009;2009:5633-6.
35. Bonato P, Sherrill DM, Standaert DG, Salles SS, Akay M. Data mining techniques to detect motor fluctuations in Parkinson's disease. *Conf Proc IEEE Eng Med Biol Soc* 2004;7:4766-9.
36. Walker PR, Smith B, Liu QY, Famili AF, Valdés JJ, Liu Z, et al. Data mining of gene expression changes in Alzheimer brain. *Artif Intell Med* 2004;31(2):137-54.
37. Miller DH, Weinshenker BG, Filippi M, Banwell BL, Cohen JA, Freedman MS, et al. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Mult Scler* 2008;14(9):1157-74.
38. Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am J Hum Genet* 2003;72(3):710-6.
39. Rumrill PD, editor. Employment Issues and Multiple Sclerosis. New York, NY: Demos Vermande, 2003.

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

40. Li Y, Schwartz CE. Data mining for response shift patterns in multiple sclerosis patients using recursive partitioning tree analysis. *Qual Life Res* 2011;20(10):1543-53.
41. Gironi M, Saresella M, Rovaris M, Vaghi M, Nemni R, Clerici M, et al. A novel data mining system points out hidden relationships between immunological markers in multiple sclerosis. *Immun Ageing* 2013;10:4933-10.
42. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol* 2005;58(6):840-6.

[Downloaded from tumj.tums.ac.ir on 2024-07-31]

TUMJ Tehran University Medical Journal, April 2017; Vol. 75, No. 1: 39-48 Original Article

Analyzing clinical symptoms in multiple sclerosis using data mining

Zahra Raeisi M.D.[1*]
Pantea Ramezannezad Ph.D.[2]
Marzieh Ahmadzade Ph.D.[1]
Shahram Tarahomi Ph.D.[3]

1- Department of Computer Engineering and IT, Shiraz University of Technology, Shiraz, Iran.
2- Department of Forensic Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
3- Department of Neurology, Jondi Shapoor University of Medical Sciences, Ahvaz, Iran.

[*] Corresponding author: Department of Computer Engineering and IT, Shiraz University of Technology, Modares Blvd., Shiraz, Iran.
Tel: +98 9130687102
E-mail: z.reisi@sutech.ac.ir

Abstract

Received: 28 Nov. 2016 Revised: 12 Apr. 2017 Accepted: 19 Apr. 2017 Available online: 20 Apr. 2017

Background: One of the today most common and incurable diseases that is associated with central neural system is 'MS' disease. Multiple sclerosis (MS) is a demyelinating disease in which the insulating covers of nerve cells in the brain and spinal cord are damaged. In this disease become apparent a wide spectrum of symptoms such as lose muscles control and their coordination and vision derangement. The goal of this research is to consider to two problems: 1- Recognition of effective clinical symptoms on MS disease and 2- Considering levels of effectiveness of age, sex and education levels factors on MS disease and association between these factors according to verity of categories of this disease.

Methods: Data mining science in medicine is worthy of attention with main application in diagnosis, therapy and prognosis, respectively high volume of collected datum. The data that were used in this article are about patients of Chaharmahal and Bakhtiari Province and collected by cure assistance. In this paper classification and association methods in software engineering field are used. Classification is a general process related to categorization, the process in which ideas and objects are recognized, differentiated, and understood. Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships.

Results: In consideration of first problem in this paper, concluded vision-clinical symptoms are the most effective symptoms and in consideration of second problem, concluded that from 584 records, women affected four times more than men. In other word 70% of MS patients with high graduate are in relapsing-remitting category and 62.5% of MS patients are 20-40 years old.

Conclusion: Some of symptoms are quite temporary and transitory and are ignored by people. Awareness of clinical-symptoms prevalence manner can be warning for people before starting critical cycle of illness. This would cause early diagnosis, effective therapy and even prevention of disease progress, respectively to MS chronicity.

Keywords: association, classification, data mining, multiple sclerosis.

[Downloaded from tumj.tums.ac.ir on 2024-07-31]