

ارزیابی سطح کنترل آسم با عبور از رویکرد مراقبتی واکنشی به یک رویکرد پیشگیرانه بر پایه خوشه‌بندی و دسته‌بندی

رقیه خشنا^۱، محمد مهدی سپهری^۲، نسرين طاهرخانی^۳

چکیده

زمینه و هدف: آسم یک بیماری مزمن غیرقابل درمان، اما قابل کنترل است که پزشکان جهت دستیابی به سطح مطلوب کنترل بیماری، نظارت مداوم بر علائم و همچنین تنظیم یک برنامه درمانی مبتنی بر خودمراقبتی را پیشنهاد می‌نمایند. ارائه این برنامه، مطابق با سطح کنترلی که بیمار در آن قرار دارد، تنظیم می‌گردد. لذا ارزیابی و دسته‌بندی دقیق سطح کنترل آسم، می‌تواند در ارائه برنامه درمانی موثر به بیمار حایز اهمیت بوده و موجب بهبود خودمراقبتی و توسعه‌ی مداخلات پیشگیرانه جهت کاهش علائم آسم شود.

روش بررسی: در این مقاله، داده‌های ۹۶ بیمار آسمی شامل اطلاعات دموگرافیک، متغیرهای بالینی و سوابق پزشکی بیمار، داده‌های زیست محیطی موثر بر آسم، معیارهای عملکرد ریه و گروه محرک مبتنی بر پرسش‌نامه‌های کنترل آسم، در یک دوره‌زمانی ۹ ماهه از یک بیمارستان تخصصی بیماری‌های ریوی تهران جمع‌آوری و در یک پایگاه‌داده چندمتغیره و چندکلاسه تجمیع شده و سپس سطح کنترل آسم با کمک یک مدل تلفیقی مبتنی بر خوشه‌بندی فازی و الگوریتم‌های با نظارت در یادگیری ماشین دسته‌بندی شده‌است.

یافته‌ها: مدل پیشنهادی برای ارزیابی سطح کنترل آسم که حاصل عملیات متوازن‌سازی، خوشه‌بندی فازی و انتخاب مشخصه بر روی داده‌هاست، دقتی به میزان ۸۸٪ ارائه نموده‌است.

نتیجه‌گیری: مدل حاضر، علاوه بر کمک به پزشکان برای شناسایی دقیق‌تر سطح کنترل آسم، می‌تواند در سیستم‌های الکترونیکی خودمراقبتی به منظور ارائه هشدارهای شخصی شده در مورد احتمال تضعیف کنترل آسم به کار رود. این چنین ابزارهایی می‌توانند مراقبت از آسم را از رویکرد مراقبتی واکنشی به یک رویکرد پیشگیرانه ارتقا دهند که در آن تصمیم‌ها و اقدام‌های درمانی از سوی پزشک ناشی از سطح کنترل آسم مزمن و پیشگیری از آسم حاد باشد.

واژه‌های کلیدی: کنترل آسم، پیشگیرانه، خوشه‌بندی، دسته‌بندی، خودمراقبتی

دریافت مقاله: اردیبهشت ۱۳۹۹
پذیرش مقاله: مرداد ۱۳۹۹

* نویسنده مسئول:

محمد مهدی سپهری؛

دانشکده مهندسی صنایع و سیستم‌ها دانشگاه
تربیت مدرس

Email :
mehdi.sepohri@modares.ac.ir

۱ دکتری مهندسی فناوری اطلاعات، قطب علمی مهندسی سیستم‌های سلامت، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

۲ استاد گروه مهندسی سیستم‌های سلامت، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران

۳ مربی گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه پیام نور، ساوه، ایران

مقدمه

آسم یک بیماری مولتی فاکتوریل، ناهمگون، دوره‌ای و تکرارپذیر است که در افراد مستعد از نظر ژنتیک و تحت تاثیر عوامل محیطی، ایجاد و یا تشدید می‌گردد و با واکنش بیش از حد مجاری هوایی به طیف وسیعی از محرک‌ها شناخته می‌شود؛ این پاسخ بیش از حد، به التهاب و انسداد مجاری هوایی می‌انجامد (۱). ماهیت ذاتی آسم و طیف وسیعی از محرک‌های تشدیدکننده آن، مدیریت این بیماری را به یک موضوع چالش برانگیز تبدیل نموده است (۲). براساس دستورالعمل‌های قدیمی‌تر توسعه یافته برای آسم، پزشکان ابتدا بیماری را بر اساس پارامتر شدت آسم (Asthma severity) طبقه‌بندی نموده و سپس فرایند درمان را تنها بر اساس همین یک فاکتور، برنامه‌ریزی می‌نمایند. این درحالی است که این نوع رویکرد در فاز اجرای گایدلاین‌ها، با نقاط ضعف متعددی همراه است (۳). مهمترین مسئله پاسخ به این پرسش است که چگونه می‌توان یک بیماری دارای ماهیت کاملا متغیر و پرنوسان مانند آسم را با چنین اندکس‌های ثابتی طبقه‌بندی نموده و درمان را تنها بر اساس شدت آن پایه‌ریزی نمود. به همین دلیل در طی سالیان اخیر، متون پزشکی توصیه می‌نمایند که تغییرات روند درمان براساس پاسخ به یک پرسش ساده‌ی دیگر باشد که آیا آسم بیمار در وضعیت کنترل است یا خیر (۳)؟

از این رو پزشکان مطابق با این رویکرد جدید، ابتدا درمان را براساس پارامتر شدت بیماری و در نظر گرفتن فاکتورهای خطر آغاز نموده و سپس در ویزیت‌های دوره‌ای از بیمار، برنامه‌ی درمانی را بر اساس سطح کنترلی که بیمار در آن قرار دارد، تنظیم مجدد می‌نمایند. در اینجا منظور از کنترل آسم، کنترل علائم تشدیدکننده‌ی بیماری است که به سه سطح کنترل شده (Well_controlled)، کنترل نشده (Not_well_controlled) و کنترل بسیار ضعیف (Very_poorly_controlled) دسته‌بندی می‌شود (۴). در نتیجه با توجه به اینکه ارزیابی پارامتر سطح کنترل آسم، به بخش لاینفک بسیاری از مداخلات درمانی این بیماری تبدیل شده است، دسته‌بندی دقیق آن برای تنظیم برنامه درمانی توسط پزشک، حیاتی و ضروری به نظر می‌رسد.

با بررسی هدفمند مقالات در حوزه‌ی ارزیابی و دسته‌بندی سطح کنترل آسم می‌توان به این نتیجه دست یافت که آنچه تحقیقات این حوزه را از هم متمایز می‌نماید در گرو پاسخ به این سه پرسش اساسی است:

(الف) هدف از بررسی سطح کنترل آسم در این تحقیقات چیست؟

(ب) کدام یک از گروه محرک‌های موثر بر سطح کنترل آسم به‌عنوان ورودی مدل‌های تحلیلی در نظر گرفته شده‌اند؟

(پ) و متناسب با هدف مطالعه چه ابزارهای تحلیلی به کار رفته و دقت و کارایی آنها به چه میزان بوده است؟

اگر از منظر هدف مطالعه (الف)، مقالات را بررسی نماییم، شاهد آن خواهیم بود که محققان بر روی دو هدف عمده تمرکز نموده‌اند: پیش‌بینی حملات آسمی به‌عنوان یک رویداد ناشی از نامطلوب‌ترین سطح کنترلی آسم در بلندمدت (۱۲-۵) و پیش‌بینی تضعیف سطح کنترل آسم (۱۳). هر دو این اهداف، رخدادهایی را به‌صورت دیر هنگام پیش‌بینی می‌نمایند که خیلی قبل‌تر از آن وقتی آسم در سطح کنترل نشده قرار داشته است، قابل شناسایی بوده‌اند. به همین جهت در این مقاله، هدف از مطالعه، دسته‌بندی سطح کنترل آسم به سه کلاس کنترل شده، کنترل نشده و کنترل بسیار ضعیف می‌باشد. با کمک این دسته‌بندی، می‌توان یک برنامه خودمراقبتی را با توجه به وضعیت بیمار و از نظر سطح کنترلی که در آن قرار دارد، به وی ارائه نمود.

از منظر دوم (ب)، می‌توان مقالات را براساس محرک‌های موثر مورد مطالعه بر پارامتر کنترل آسم، از هم متمایز نمود. این محرک‌ها را می‌توان به هشت دسته مختلف تقسیم‌بندی نمود که هر کدام از آنها خود شامل زیر محرک‌های متنوعی هستند که همه یا بخشی از آنها در مقالات گذشته مورد مطالعه قرار گرفته‌اند. این گروه محرک‌ها عبارتند از (۱۴):

- گروه محرک مبتنی بر اطلاعات دموگرافیک بیمار
- گروه محرک مبتنی بر متغیرهای بالینی و سوابق پزشکی بیمار
- گروه محرک مبتنی بر داده‌های زیست‌محیطی
- گروه محرک مبتنی بر معیارهای عملکرد ریه (Lung Function)
- گروه محرک مبتنی بر عوامل ژنتیکی

● گروه محرک مبتنی بر نشانگرهای زیستی

● گروه محرک مبتنی بر پرسش‌نامه‌ها و گایدلاین‌های کنترل آسم از منظر ابزار تحلیل داده‌ها(پ)، محققان از تکنیک‌های مختلفی شامل انواع روش‌های آماری و ابزارهای داده‌کاوی استفاده نموده‌اند. هرکدام از این روش‌ها، مزایا و معایب مخصوص به خود را دارند و محققان بسته به آنچه هدف پژوهش آنان است، یک یا تعدادی از این روش‌ها را برمی‌گزینند. سادگی استفاده، میزان قدرت محاسباتی، قابلیت تحلیل نتایج توسط کاربران، میزان دقت الگوریتم از جمله مواردی است که در هنگام انتخاب یک مدل مناسب مدنظر قرار می‌گیرند. در این میان، استفاده از تحلیل‌های داده‌کاوی روش محبوب بسیاری از پژوهشگران پیشین می‌باشد. داده‌کاوی به معنای استخراج اطلاعات نهان، الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ است(۱۰). الگوریتم‌های پرکاربرد در مباحث مربوط به کنترل آسم، شامل الگوریتم‌های دسته‌بندی (Classification Algorithm) نظیر رگرسیون لجستیک (Logistic regression)(۷و۱۵)، ماشین‌بردار پشتیبان (Support vector machine)(۸و۱۳)، الگوریتم بیزین (Bayesian method)(۸و۱۳و۱۶)، درخت تصمیم (Decision tree)(۱۰و۱۶)، الگوریتم جنگل تصادفی(۱۱) و نزدیکترین درخت همسایگی (K-nearest neighbor)(۱۳) می‌باشد. در به‌کارگیری این روش‌ها، آنچه

حایز اهمیت است، ارزیابی عملکرد مدل‌هاست. نتایج مقالات پیشین نشان می‌دهند که عملکرد مدل‌ها از یک پایگاه داده به یک پایگاه داده‌ی دیگر با هم متفاوت است و بستگی به مدل انتخابی پژوهشگران دارد و نمی‌توان مدل واحدی را به‌عنوان بهترین مدل برگزید. به همین جهت در این مطالعه، طیف متنوعی از دسته‌بندها را بر روی پایگاه داده اعمال نموده و با ارزیابی نتایج، بهترین مدل را برمی‌گزینیم؛ زیرا برای ساخت یک مدل دسته‌بندی قدرتمند، لازم است ترکیبی مناسب از مدل‌ها و ترکیب مناسبی از محرک‌های اثرگذار بر سطح کنترل آسم انتخاب شوند.

لذا با توجه به اهمیت موضوع، در این مطالعه قصد داریم تا با کمک الگوریتم‌های متنوع دسته‌بندی در داده‌کاوی، سطح کنترل آسم را ارزیابی نماییم. دسته‌بندی دقیق سطح کنترل آسم می‌تواند منجر به پیش‌آگاهی زود هنگام در مورد تضعیف کنترل آسم شده و در مواقعی که بیمار در وضعیت نامناسبی قرار دارد، می‌توان به وی و پزشک معالجش جهت اقدامات عاجل، هشدارهای لازم را ارسال نمود. همچنین شناسایی افراد در معرض تضعیف کنترل آسم، می‌تواند به مدیریت موثرتر بیماری کمک نموده و در نهایت درد، رنج، مرگ و هزینه‌های بهداشتی و درمانی حاصل از آن را کاهش داد. از سویی دیگر، با شناسایی سطحی که کنترل بیماری در وضعیت مناسبی قرار دارد، پزشک می‌تواند با ادامه‌ی برنامه‌ی درمانی پیشین خود، وضعیت مطلوب بیمار را حفظ و نگهداری نماید.

روش بررسی

۱. جمع‌آوری و تجمیع داده‌های موثر بر سطح کنترل آسم در یک پایگاه داده چندمتغیره

۲. آماده‌سازی و پیش‌پردازش داده‌های چندمتغیره موثر بر سطح کنترل آسم

۳. ایجاد و ارزیابی الگوریتم‌های دسته‌بندی باناظر برای شناسایی سطح کنترل آسم

شکل ۱: روش اجرای پژوهش

دوم، عملیات آماده‌سازی و پیش‌پردازش شامل مواجهه با داده‌های پرت و مفقوده، متوازن‌سازی داده‌ها، خوشه‌بندی و کاهش بعد و انتخاب مشخصه صورت می‌پذیرد. این مرحله مهمترین مرحله از انجام تحلیل‌های داده‌کاوی

شکل ۱ گام‌های این پژوهش را به ترتیب نشان می‌دهد. در گام ابتدایی، داده‌های مورد نیاز از منابع داده‌ای مختلف در بیمارستان جمع‌آوری شده و در یک پایگاه داده چندمتغیره تجمیع می‌شوند. در گام

است. سپس در گام پایانی، یک مدل دسته‌بندی سطح کنترل آسم با کمک الگوریتم‌های دسته‌بندی با ناظر، ارائه می‌گردد.

• جمع‌آوری داده‌ها

داده‌های این پژوهش از اطلاعات ۹۶ بیمارآسمی، ساکن مناطق مختلف شهر تهران که به درمانگاه آسم و آلرژی بزرگترین مرکز ارائه خدمات درمانی ایران در حوزه‌ی بیماری‌های تنفسی در یک بازه زمانی نه ماهه مراجعه نموده‌بودند، گردآوری شده است. اطلاعات اخذ شده از هر بیمار ۵ گروه محرک مختلف را در برمی‌گیرد که شامل محرک‌های مبتنی بر داده‌های دموگرافیک بیمار، محرک‌های مبتنی بر متغیرهای بالینی و سوابق پزشکی بیمار، محرک‌های مبتنی بر داده‌های زیست‌محیطی، محرک‌های مبتنی بر معیارهای عملکرد ریه و محرک‌های استخراج شده از پرسش‌نامه کنترل آسم (Asthma control test = ACT) می‌باشد. داده‌های دو گروه محرک‌های عوامل ژنتیکی و نشان‌گرهای زیستی در دسترس پژوهشگران این پژوهش نبوده و از مطالعه حذف شده‌اند.

برای گردآوری این مجموعه اطلاعات، در روز مراجعه‌ی بیمار به درمانگاه با کمک یک فرم ثبت استاندارد، متغیرها و پارامترهای دو گروه داده‌های دموگرافیک و سوابق پزشکی بیمار جمع‌آوری گردید. اطلاعات تست پوستی آلرژی نیز از بیماران اخذ و به این گروه از اطلاعات افزوده شد. همچنین با انجام تست اسپیرومتری، متغیرهای عملکرد ریوی بیمار نیز در پایگاه داده جمع‌آوری شد.

همچنین از بیماران مراجعه‌کننده خواسته شده تا با کمک یک دستگاه پیک‌فلومتر که متناسب با سن هر بیمار در اختیار وی قرار گرفته و نحوه‌ی کار با آن آموزش داده شده، به مدت ۴ تا ۱۲ هفته متوالی میزان شاخص حداکثر سرعت جریان بازدمی (Peak expiratory flow rate) را در دو نوبت صبح و عصر اندازه‌گیری نموده و ثبت نمایند. این شاخص، حداکثر میزان جریان هوایی را که با قدرت می‌توان از ریه‌ها خارج ساخت، اندازه‌گیری می‌کند و به‌طور گسترده‌ای برای ارزیابی و پایش روزانه‌ی کنترل آسم و همچنین پیش‌بینی وقوع حملات آسمی استفاده می‌شود. همچنین همزمان با اندازه‌گیری این شاخص، سایر علائم آسم شامل بروز تنگی نفس، خس‌خس، سرفه، تنگی قفسه‌سینه، درد، محدودیت در فعالیت‌های روزانه و

بیداری شبانه و نیز اطلاعات مربوط به نیاز به مصرف داروهای استنشاقی نیز توسط بیماران جمع‌آوری گردیده است. همچنین با کمک این اطلاعات، ACT score متعلق به هر بیمار در بازه‌های هفتگی محاسبه و به‌عنوان متغیر مستقل در نظر گرفته شده است. نمره‌ی ۲۵ به معنای سطح کنترل شده، نمره‌ی ۲۰ تا ۲۴ به معنای سطح کنترل نشده، و نمره‌ی کمتر از ۲۰ به معنای سطح کنترل بسیار ضعیف امتیازگذاری می‌شود (۱۷).

برای تکمیل پایگاه داده علاوه بر موارد فوق به محرک‌های زیست‌محیطی شامل پارامترهای آلودگی هوا و داده‌های آب و هواشناسی به‌طور روزانه و برای هر رکورد داده با توجه به تاریخ و آدرس محل زندگی بیماران، نیاز است. اطلاعات آلاینده‌های هوا شامل CO, O3, NO2, SO2, PM10, و PM2.5, AQI و وضعیت کیفیت هوا در روز مربوط از پایگاه داده‌ی ۲۴ ایستگاه سنجش آلاینده‌ی شرکت کنترل آلودگی هوای تهران اخذ گردیده است. همچنین اطلاعات مرتبط با پارامترهای آب و هواشناسی شامل رطوبت نسبی، متوسط دمای هوا، حداکثر و حداقل درجه حرارات مطلق، متوسط میزان وزش باد، فشارنسبی و میزان بارندگی، نقطه شبنم و نوع رخداد جوی از اطلاعات چهار ایستگاه سینوپتیک سازمان هواشناسی در این شهر استخراج شده و به هر رکورد داده‌ی بیمار با کمک ابزار Spatial Join نرم افزار ArcGIS 10.3 متصل شده است.

حال به‌منظور انجام تحلیل‌های داده‌کاوی، پنج پایگاه داده‌ی فوق را در یک پایگاه داده‌ی چندمتغیره تجمیع نموده که شامل یک متغیر پاسخ سه کلاسه با نام منطقه کنترلی آسم شامل سه کلاس آسم کنترل شده یا منطقه‌ی کنترلی سبز رنگ (Green zone)، آسم کنترل نشده یا منطقه‌ی کنترلی زرد رنگ (Yellow zone) و کنترل بسیار ضعیف یا منطقه‌ی کنترلی قرمز رنگ (Red zone) می‌باشد. پایگاه داده تجمیعی شامل ۲۸۷۰ رکورد داده و ۱۴۳ متغیر مستقل است.

• پیش‌پردازش داده‌ها

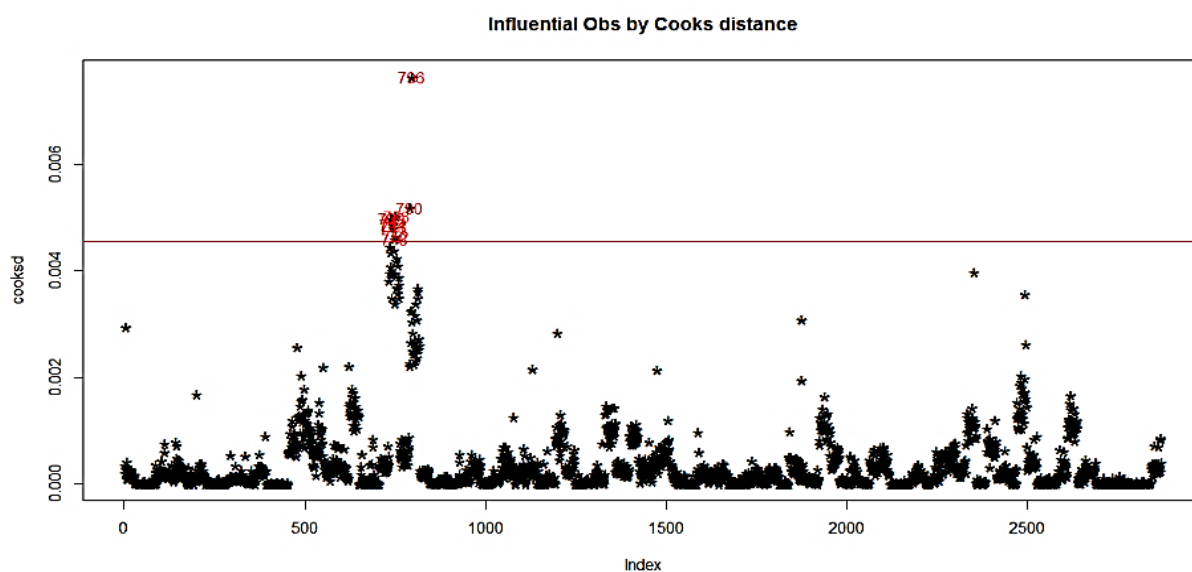
پیش‌پردازش صحیح داده‌ها، هم داده را آماده و هم داده‌کاوی را تجهیز می‌نماید و این به معنی آماده‌سازی داده‌ها به شکلی قابل استفاده در مدل و ساخت صحیح مدل است. از این رو بسته به داده‌ها و مدل انتخابی،

ویژگی‌ها و هم بر روی مشاهدات بررسی می‌شوند. به همین منظور، به‌طور جداگانه برای تمامی متغیرهای پیوسته نمودار جعبه‌ای رسم شده و بیشترین داده‌های پرت نیز مربوط به گروه متغیرهای زیست‌محیطی هستند. همچنین، برای شناسایی مشاهدات پرت، از یک رویکرد چندمتغیره به نام Cook's distance (۱۸) که مبتنی بر یک مدل رگرسیونی است، استفاده می‌کنیم.

ممکن است تکنیک‌های مختلفی مورد استفاده قرار گیرند. در این مقاله، عملیات شناسایی داده‌های پرت، تخمین مقادیر مفقوده، نرمال‌سازی داده‌ها به روش نمره استاندارد (Z-score)، خوشه‌بندی فازی، انتخاب ویژگی و نمونه‌برداری مجدد برای مواجهه با داده‌هایی با کلاس‌های نامتوازن صورت می‌پذیرد.

● شناسایی داده‌های پرت

در اولین مرحله از آماده‌سازی، وجود داده‌های پرت هم بر روی



شکل ۲: بررسی مشاهدات پرت با رویکرد Cook's distance

(۱۹) Multiple imputation using Fully Conditional Specification استفاده نموده‌ایم. همچنین برای سایر گروه متغیرها، میزان ارزش‌های از دست رفته، کمتر از ۲٪ بوده و با کمک روش میانگین مشاهدات، خانه‌های خالی را پر می‌نماییم. بعد از انجام این مرحله، پایگاه داده‌ی مورد مطالعه، دارای ۱۴۰ متغیر و ۲۸۷۰ مشاهده است.

● مواجهه با داده‌هایی با کلاس نامتوازن

وجود کلاس‌های نامتوازن در داده‌ها موجب کاهش کارایی تکنیک‌های دسته‌بندی می‌شوند.

مشاهداتی که بالای خط قرمز قرار می‌گیرند مشاهدات پرت هستند و پس از بررسی بیشتر در صورت لزوم از مجموعه داده حذف شده‌اند.

● تخمین مقادیر مفقود

بیشترین ارزش‌های از دست رفته متعلق به گروه متغیرهای عملکرد ریوی بوده و میزان آن نیز قابل توجه است؛ لذا مناسب نیست که خانه‌های خالی با روش ساده‌ای همچون میانگین‌گیری پر شوند، زیرا موجب کاهش واریانس داده‌ها خواهد شد که نامطلوب است. به همین جهت از روش

جدول ۱: مجموعه داده‌های اصلی و متوازن‌شده

| مجموعه داده | اندازه داده | داده‌های با کلاس سبز | داده‌های با کلاس زرد | داده‌های با کلاس قرمز |
|----------------------------|-------------|----------------------|----------------------|-----------------------|
| مجموعه داده‌ی شماره ۱ (D۱) | ۲۸۷۰ | ۱۸۷۸ | ۹۵۴ | ۳۸ |
| | | (/۶۵/۴۳) | (/۳۳/۲۴) | (/۱/۳۲) |

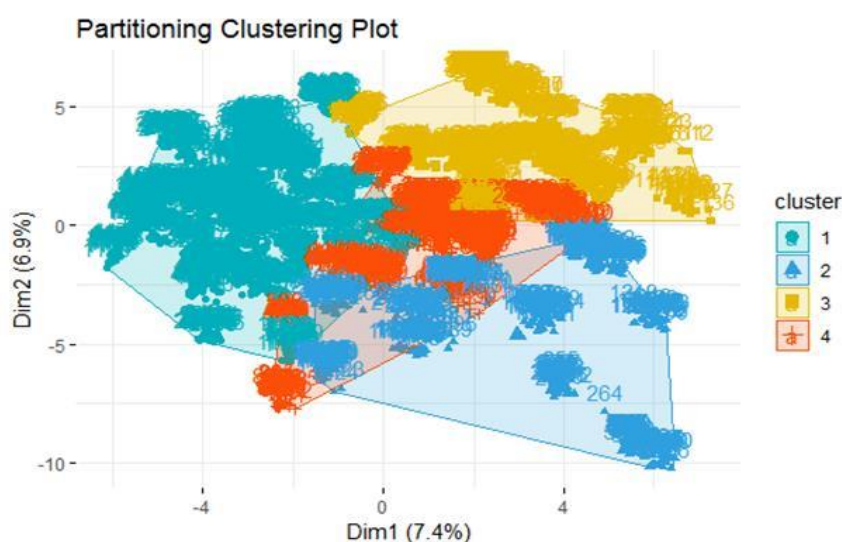
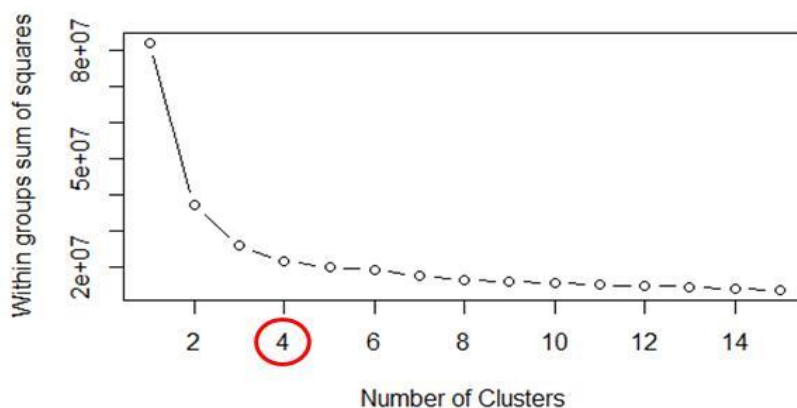
| | | | | |
|------------|------------|------------|------|-------------------------------|
| ۱۳۱۵ | ۱۳۱۵ | ۱۳۱۵ | ۳۹۴۵ | مجموعه داده‌ی شماره ۲ (D۲) |
| (/۰.۳۳/۳۳) | (/۰.۳۳/۳۳) | (/۰.۳۳/۳۳) | | |
| ۶۹۹ | ۶۹۹ | ۶۷۰ | ۲۰۶۸ | مجموعه داده‌ی شماره ۳ (D۳) |
| (/۰.۳۳/۸۰) | (/۰.۳۳/۸۰) | (/۰.۳۲/۳۹) | | |

همان‌طور که از جدول ۱ مشخص است در مجموعه داده‌ی اولیه، نمونه‌های متعلق به کلاس سبز ۶۵/۴۳٪، تعداد نمونه‌های متعلق به کلاس زرد ۳۳/۲۴٪ و داده‌های متعلق به کلاس قرمز ۱/۳۲٪ داده‌ها را پوشش می‌دهند. این دسته‌بندی نشان از نامتوازن بودن کلاس‌ها دارد. برای حل این چالش، ابتدا داده‌ها را به نسبت ۷ به ۳، به دو مجموعه داده‌ی آموزش و آزمون تقسیم‌بندی نموده و سپس مجموع داده‌ی شماره ۲ را با یک روش بیش نمونه‌برداری و مجموع داده شماره ۳ را با روش باز نمونه‌گیری

بر روی مجموعه داده‌ی شماره ۱ (Synthetic Minority Over-sampling Technique) SMOTE (۲۰) فرایند خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت و فرایند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگرند، تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود.

• خوشه‌بندی فازی بر روی مجموعه داده‌ها

خوشه‌بندی فازی بر روی مجموعه داده‌ها



شکل ۳: تعداد بهینه خوشه‌ها و نتایج خوشه‌بندی فازی بر روی داده‌های مورد مطالعه

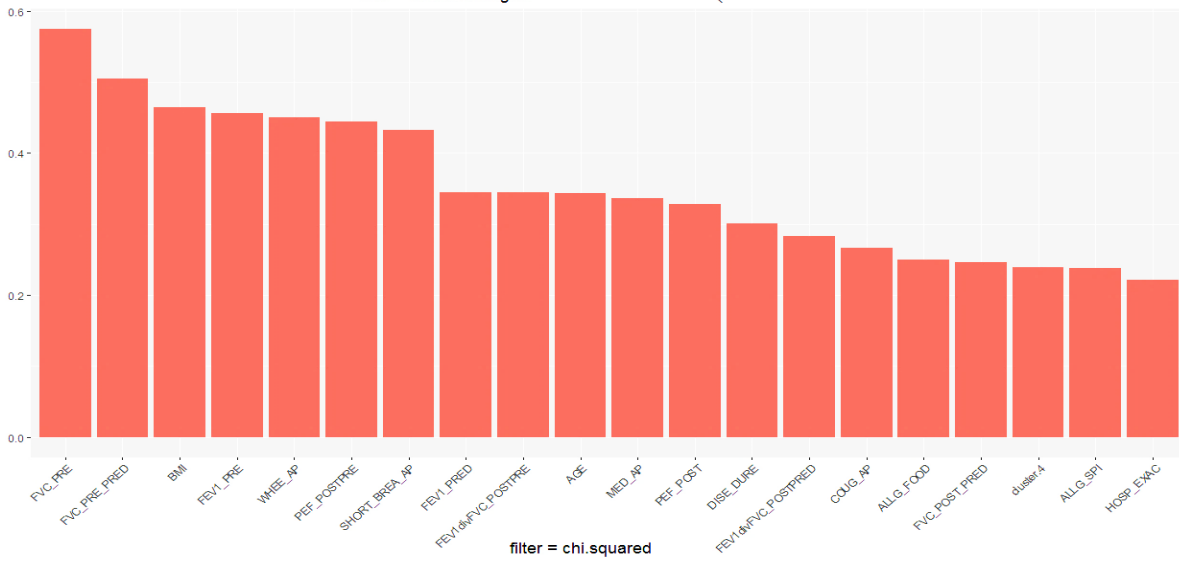
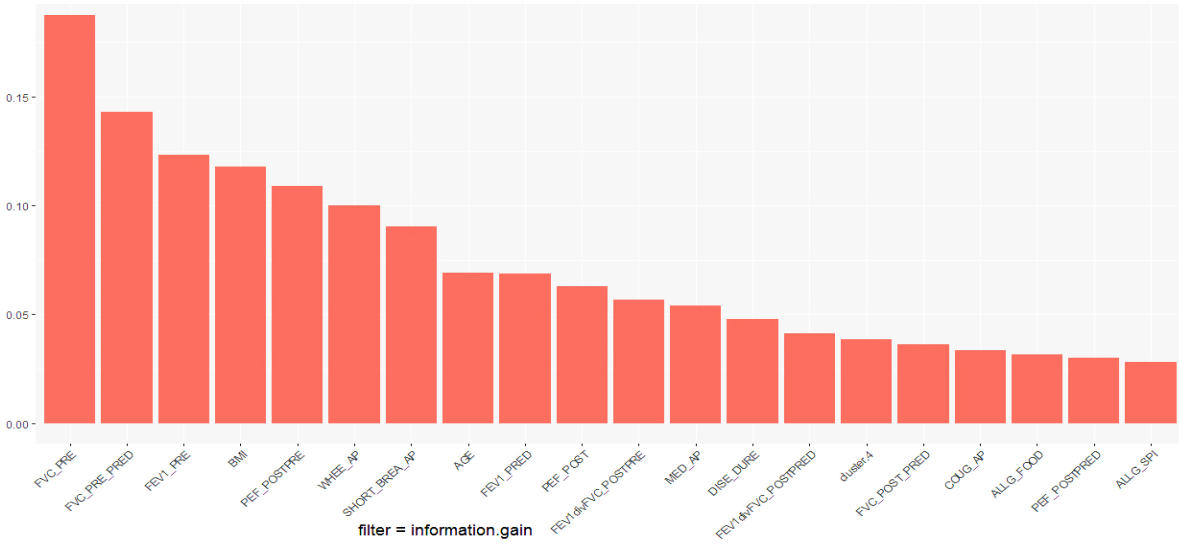
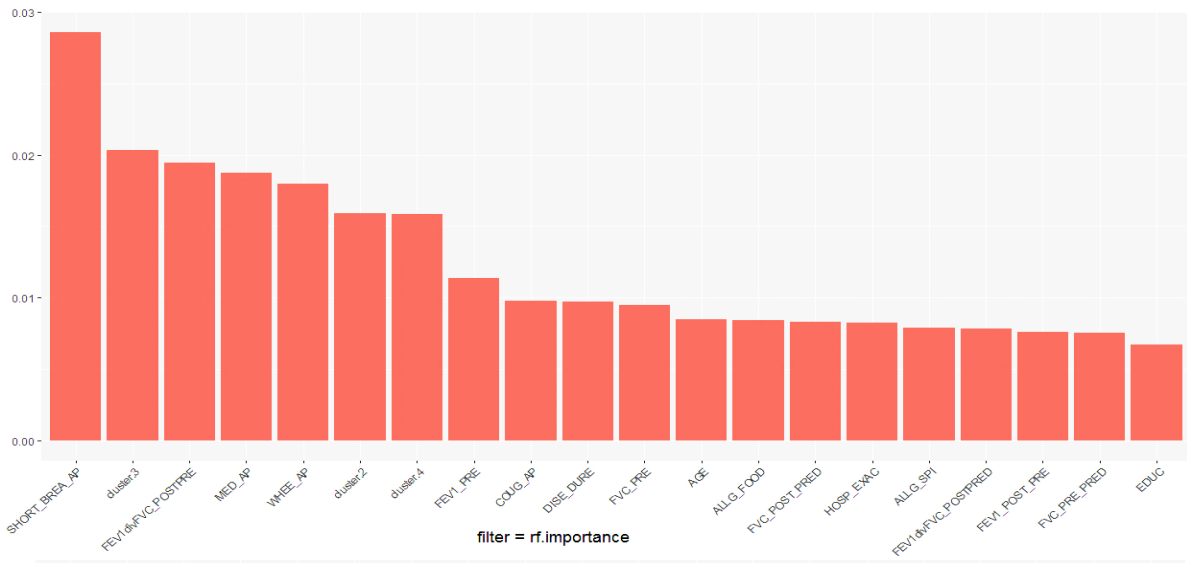
در این مطالعه هدف از خوشه‌بندی، اضافه نمودن شماره خوشه‌ها به مجموعه ویژگی‌هاست. در بسیاری از موارد، این عمل موجب بهبود قابل توجه در دقت نهایی مدل دسته‌بندی خواهد شد. به همین منظور، از الگوریتم c_mean که یک الگوریتم خوشه‌بندی فازی است، استفاده می‌کنیم و نتایج حاصل از آن را به‌عنوان متغیر جدید به مجموعه داده‌ها می‌افزاییم. مطابق شکل ۳، تعداد بهینه خوشه‌ها، چهار می‌باشد.

در این مطالعه هدف از خوشه‌بندی، اضافه نمودن شماره خوشه‌ها به مجموعه ویژگی‌هاست. در بسیاری از موارد، این عمل موجب بهبود قابل توجه در دقت نهایی مدل دسته‌بندی خواهد شد. به همین منظور، از الگوریتم c_mean که یک الگوریتم خوشه‌بندی فازی است، استفاده می‌کنیم و نتایج حاصل از آن را به‌عنوان متغیر جدید به مجموعه داده‌ها می‌افزاییم. مطابق شکل ۳، تعداد بهینه خوشه‌ها، چهار می‌باشد.

● انتخاب مشخصه

افزایش کارایی مدل‌های دسته‌بندی است.

هدف از انتخاب مشخصه، کاهش بعد مجموعه داده به منظور



شکل ۱۴: فرایند انتخاب مشخصه بر روی داده‌های مورد مطالعه

یافته‌ها

در این قسمت مدل‌های مختلف برای دسته‌بندی سطح کنترل آسم را ایجاد و ارزیابی می‌کنیم. در مطالعات پیشین هیچ الگوریتم مشخصی را نمی‌توان یافت که برای پیش‌بینی و دسته‌بندی در حوزه بیماری آسم برای تمامی مسایل و در تمامی زمانها بهترین (دقیق‌ترین) یادگیرنده و دسته‌بند را به وجود آورد. به همین جهت، در این مقاله از الگوریتم‌های مختلفی همچون رگرسیون، *xgboost*، جنگل تصادفی، درخت تصمیم، دسته‌بند بیزین گاوسی و ماشین بردار پشتیبان استفاده نموده‌ایم. روش‌های انتخابی از بین روش‌های پرکاربرد مقالات پیشین برای این مسئله انتخاب شده‌اند. سپس با ارزیابی نتایج به دست آمده، بهترین مدل را بر می‌گزینیم. شاخص‌های ارزیابی عملکرد به کار گرفته شده برای مقایسه‌ی نتایج عبارتند از: صحت (*Accuracy*)، دقت (*Precision*)، حساسیت (*Sensitivity*) و ویژگی (*Specificity*). قابل ذکر است که همه‌ی مدل‌ها با کمک نرم‌افزار *Rstudio 1.1.453* کدنویسی شده‌اند.

دلایل انتخاب این شاخص‌های ارزیابی به این علت است که در بحث دسته‌بندی سطح کنترل آسم بسیار مهم است که کلاس‌های دارای سطح کنترل نامطلوب و خیلی نامطلوب (ناحیه زرد و قرمز) درست تشخیص داده‌شوند تا کلاس‌های دارای وضعیت مطلوب (ناحیه سبز)؛ زیرا اگر کلاس منطقه‌ی کنترلی قرمز به اشتباه زرد یا سبز تشخیص داده‌شود و یا منطق کنترلی زرد، سبز دسته‌بندی شود، ممکن است به علت ارایه برنامه درمانی نامناسب و یا ناکافی، بیمار دچار وضعیت کنترلی وخیم و در نهایت حملات آسمی گردد که هزینه‌های به مراتب بیشتری برای بیماران و سیستم‌های درمانی در پی دارد. در مبحث بررسی شاخص‌ها، این نکته قابل ذکر است که در نظر گرفتن شاخص صحت، به تنهایی برای سنجش کارایی یک مدل کافی نیست؛ زیرا در مجموعه داده‌هایی با مشکل عدم توازن کلاس‌ها، نتایج، فریبنده خواهند بود. به همین جهت در اینگونه مسایل، بهبود میزان دو شاخص حساسیت و صحت در کنار هم، به دلیل کشف

در این پژوهش، از سه روش فیلتر شامل جنگل تصادفی با شاخص (*mean decrease accuracy*(MDA)، یک روش مبتنی بر آنتروپی با شاخص *information gain* یا *IG* و همچنین تست آماری *Pearson's chi-squared* استفاده می‌نماییم. فیلترها بر ویژگی‌های کلی مجموعه داده‌ی آموزش تکیه دارند و فرایند انتخاب ویژگی را به عنوان یک گام پیش‌پردازش با استقلال از الگوریتم استقرایی انجام می‌دهند. مزیت این مدل‌ها هزینه‌ی محاسباتی پایین و توانایی تعمیم خوب آن‌هاست. شاخص *MDA* به‌طور مستقیم تاثیر هر یک از ویژگی‌ها را بر دقت مدل می‌سنجد. ایده‌ی کلی این است که مقادیر هر یک از ویژگی‌ها را تغییر دهیم و میزان تغییر در دقت مدل را اندازه‌گیری کنیم. واضح است که برای متغیرهای بی‌اهمیت، جایگزینی باید تا حدودی بر دقت مدل اثر نگذارد، درحالی‌که تغییر متغیرهای مهم باید دقت را به‌طور قابل توجهی کاهش دهد. *IG* میزان شاخص آنتروپی را اندازه‌گیری کرده و نشان می‌دهد که کدام دسته از ویژگی‌ها در مجموعه داده‌ی آموزشی، بهتر می‌تواند برچسب کلاس‌های یادگیری را از هم متمایز و جدا نمایند. همچنین شاخص کای پیرسون ارزش متغیرها را با آماره‌ی *chi-squared* با توجه به هر برچسب کلاس یادگیری می‌سنجد. اگر متغیر هدف مستقل آماری از متغیر پیش‌بینی کننده تشخیص داده شود، می‌توان این متغیر را از مجموعه داده‌ها حذف کرد.

با توضیحات فوق، هر کدام از این سه روش را به‌طور جداگانه بر روی ۱۴۳ متغیر اصلی اعمال و متغیرهای مهم‌تر در همه روش‌ها را با هم تلفیق می‌نماییم. دلیل استفاده از این رویکرد تلفیقی این است که هیچ متغیر با اهمیتی را از دست ندهیم و پایگاه داده شامل با اهمیت‌ترین متغیرها باشد. همچنین بعد از کاهش تعداد متغیرهای مستقل، سرعت الگوریتم‌های دسته‌بندی بهبود یابد. بعد از مرحله‌ی انتخاب مشخصه، پایگاه داده‌ی کاهش یافته شامل ۲۸۷۰ مشاهده و ۳۵ متغیر است. شکل اهمیت این متغیرها را با توجه به شاخص‌های مربوط نمایش می‌دهد.

سطح کنترل آسم در دو وضعیت زرد و قرمز اهمیت بیشتری دارد.

جدول ۲: ارزیابی عملکرد مدل‌های مورد استفاده برای دسته‌بندی سطح کنترل آسم

| بیزین | | | بوستینگ | | | رگرسیون لجستیک چندگانه | | | ارزیابی عملکرد | |
|------------|-------|-------|-------------|-------|-------|------------------------|-------|-------|----------------|--------|
| D۳ | D۲ | D۱ | D۳ | D۲ | D۱ | D۳ | D۲ | D۱ | | |
| ۰/۸۶۷ | ۰/۸۶۹ | ۰/۸۳۵ | ۰/۹۴۲ | ۰/۹۲۱ | ۰/۹۳۴ | ۰/۹۲۷ | ۰/۹۲۵ | ۰/۹۱۲ | Green_Zone | |
| ۰/۶۰۳ | ۰/۵۹۹ | ۰/۶۳۰ | ۰/۸۳۵ | ۰/۸۸۱ | ۰/۸۶۲ | ۰/۷۹۰ | ۰/۷۸۹ | ۰/۸۴۹ | Yellow_Zone | حساسیت |
| ۰/۱۳۸ | ۰/۱۶۱ | ۰/۲۱۶ | ۰/۷۴۸ | ۰/۸۳۵ | ۰/۶۳۷ | ۰/۷۲۰ | ۰/۶۶۲ | ۰/۷۷۰ | Red_Zone | |
| ۰/۶۳۰ | ۰/۶۳۵ | ۰/۶۹۲ | ۰/۸۵۱ | ۰/۸۸۸ | ۰/۸۷۷ | ۰/۸۰۳ | ۰/۷۸۲ | ۰/۸۵۲ | Green_Zone | |
| ۰/۸۰۶ | ۰/۸۱۰ | ۰/۸۱۶ | ۰/۹۳۹ | ۰/۹۲۰ | ۰/۹۳۱ | ۰/۹۲۲ | ۰/۹۱۹ | ۰/۹۱۲ | Yellow_Zone | ویژگی |
| ۰/۹۵۶ | ۰/۹۵۸ | ۰/۹۵۶ | ۰/۹۵۵ | ۰/۹۵۵ | ۰/۹۵۲ | ۰/۹۵۸ | ۰/۹۵۸ | ۰/۹۵۶ | Red_Zone | |
| ۰/۸۳۴ | ۰/۷۱۷ | ۰/۷۷۷ | ۰/۹۰۳ | ۰/۹۴۳ | ۰/۹۱۶ | ۰/۸۶۲ | ۰/۸۶۲ | ۰/۹۰۳ | Green_Zone | |
| ۰/۶۵۹ | ۰/۶۷۳ | ۰/۶۶۶ | ۰/۹۱۱ | ۰/۸۶۹ | ۰/۸۹۳ | ۰/۸۸۹ | ۰/۸۸۲ | ۰/۸۵۵ | Yellow_Zone | دقت |
| ۰/۷۰۷ | ۰/۸۰۸ | ۰/۷۰۷ | ۰/۶۳۶ | ۰/۶۳۶ | ۰/۳۶۴ | ۰/۸۰۸ | ۰/۸۰۸ | ۰/۷۰۷ | Red_Zone | |
| ۰/۷۷۹ | ۰/۷۷۲ | ۰/۷۸۳ | ۰/۸۶۷ | ۰/۸۹۹ | ۰/۹۴۱ | ۰/۸۵۵ | ۰/۸۵۴ | ۰/۹۱۷ | Green_Zone | |
| ۰/۸۳۰ | ۰/۷۲۴ | ۰/۷۵۸ | ۰/۸۸۷ | ۰/۹۰۱ | ۰/۹۲۶ | ۰/۸۳۶ | ۰/۸۳۴ | ۰/۹۱۰ | Yellow_Zone | صحت |
| ۰/۵۶۷ | ۰/۵۷۹ | ۰/۶۰۶ | ۰/۸۵۷ | ۰/۸۹۵ | ۰/۸۲۹ | ۰/۸۵۴ | ۰/۸۱۵ | ۰/۹۹۸ | Red_Zone | |
| درخت تصمیم | | | جنگل تصادفی | | | ماشین بردار پشتیبان | | | ارزیابی عملکرد | |
| D۳ | D۲ | D۱ | D۳ | D۲ | D۱ | D۳ | D۲ | D۱ | | |
| ۰/۸۸۰ | ۰/۸۷۸ | ۰/۸۵۹ | ۰/۹۰۹ | ۰/۹۱۷ | ۰/۹۱۵ | ۰/۹۰۷ | ۰/۸۹۰ | ۰/۹۳۴ | Green_Zone | |
| ۰/۷۵۴ | ۰/۸۲۸ | ۰/۸۱۸ | ۰/۸۲۹ | ۰/۸۷۳ | ۰/۸۸۲ | ۰/۷۱۱ | ۰/۷۷۵ | ۰/۸۵۲ | Yellow_Zone | حساسیت |
| ۰/۲۸۶ | ۰/۶۶۷ | ۰/۳۰۰ | ۰/۸۰۰ | ۰/۸۵۵ | ۰/۷۹۰ | ۰/۲۱۱ | ۰/۲۷۳ | ۰/۶۶۷ | Red_Zone | |
| ۰/۷۵۷ | ۰/۸۸۰ | ۰/۸۷۲ | ۰/۸۷۲ | ۰/۹۱۹ | ۰/۹۱۱ | ۰/۷۲۸ | ۰/۷۸۰ | ۰/۸۹۷ | Green_Zone | |
| ۰/۹۰۴ | ۰/۹۲۴ | ۰/۸۷۱ | ۰/۹۳۵ | ۰/۹۲۶ | ۰/۹۱۵ | ۰/۸۷۴ | ۰/۸۸۷ | ۰/۹۶۱ | Yellow_Zone | ویژگی |
| ۰/۹۹۶ | ۰/۹۹۴ | ۰/۹۹۱ | ۰/۹۹۶ | ۰/۹۹۵ | ۰/۹۹۴ | ۰/۹۹۶ | ۰/۹۹۸ | ۰/۹۹۲ | Red_Zone | |
| ۰/۸۲۴ | ۰/۸۸۸ | ۰/۹۱۱ | ۰/۹۰۷ | ۰/۹۴۷ | ۰/۹۶۰ | ۰/۸۱۵ | ۰/۸۵۶ | ۰/۹۲۶ | Green_Zone | |
| ۰/۸۱۵ | ۰/۸۴۶ | ۰/۷۲۴ | ۰/۸۸۴ | ۰/۹۱۳ | ۰/۹۰۹ | ۰/۷۵۵ | ۰/۷۷۳ | ۰/۸۸۳ | Yellow_Zone | دقت |
| ۰/۷۲۷ | ۰/۵۴۵ | ۰/۲۷۳ | ۰/۷۲۷ | ۰/۶۳۶ | ۰/۵۴۵ | ۰/۷۲۷ | ۰/۸۱۸ | ۰/۳۶۴ | Red_Zone | |
| ۰/۷۵۸ | ۰/۸۲۴ | ۰/۸۰۵ | ۰/۸۷۴ | ۰/۸۸۸ | ۰/۸۹۸ | ۰/۷۶۸ | ۰/۸۱۵ | ۰/۸۸۱ | Green_Zone | |
| ۰/۷۸۹ | ۰/۸۵۱ | ۰/۸۱۵ | ۰/۸۳۲ | ۰/۸۴۰ | ۰/۸۳۹ | ۰/۷۲۲ | ۰/۷۶۱ | ۰/۸۴۶ | Yellow_Zone | صحت |
| ۰/۶۴۱ | ۰/۸۱۰ | ۰/۶۴۵ | ۰/۸۶۸ | ۰/۸۸۵ | ۰/۹۵۷ | ۰/۵۷۳ | ۰/۶۰۵ | ۰/۷۹۹ | Red_Zone | |

جدول ۲، نتایج پیاده‌سازی الگوریتم‌های دسته‌بندی را نشان می‌دهد. همان‌طور که مشاهده می‌شود، تمامی مدل‌ها در دسته‌بندی کلاس سبز در هر سه مجموعه داده، نتایج تقریباً مشابه و قابل قبولی داشته‌اند. دسته‌بند بیزین برای دسته‌بندی کلاس زرد عملکرد ضعیفی داشته، در حالی که الگوریتم جنگل تصادفی و xgboost نتایج قابل قبولی ارائه داده‌اند. دسته‌بندهای رگرسیون و جنگل تصادفی و

xgboost برای شناسایی کلاس قرمز خوب عمل کرده و شاخص‌های دقت و صحت قابل‌قبولی دارند؛ در حالی که نتایج این بخش در سایر الگوریتم‌ها نامناسب و در الگوریتم‌های ماشین بردار پشتیبان و بیزین بسیار ناامیدکننده است. معیار دقت برای مجموعه داده D1 برای شناسایی کلاس قرمز در الگوریتم‌های Xgboost و ماشین بردار پشتیبان، جنگل تصادفی و درخت تصمیم نتایج خوبی نداشته و در

صورت استفاده هشدارهای اشتباه برای وقوع این کلاس افزایش خواهند یافت.

از سوی دیگر، نتایج جدول ۲ نشان می‌دهند که مدل‌های مختلف در مجموعه داده‌های تعریف شده به‌ویژه برای کلاس قرمز متفاوت عمل نموده‌اند. در این میان، اغلب معیارها برای مجموعه داده‌ی ۱، که یک مجموعه داده‌ی نامتوازن است، نتایج بهتری از D3 داشت که برخلاف انتظار ما بود. دلیل آن ممکن است به علت روش نمونه‌گیری انتخابی باشد. در حالت کلی، الگوریتم‌ها بر روی D2 بهتر از D1 و D3 و همچنین بر روی D1 نیز بهتر از D3 عمل نموده‌اند. نتایج جدول ۲ نشان می‌دهد که الگوریتم‌های جنگل تصادفی و xgboost بر روی D2 که از روش بیش نمونه‌برداری متوازن شده‌است، در دسته‌بندی هر سه کلاس بهتر از سایر روش‌ها عمل نموده‌اند. همچنین نتایج ارزیابی تجمیعی مدل‌های دسته‌بندی کنترل آسم نشان می‌دهد که الگوریتم xgboost، که یک روش یادگیری گروهی است، نتایج بهتری نسبت به سایر الگوریتم‌ها ارائه داده است و شاخص صحت آن در D2 ۰/۸۸ برآورد شده است. در جدول ۲ برای وضوح بیشتر، مقادیر با تفکیک رنگ از هم متمایز شده‌اند و رنگ قرمز پررنگ در یک سلول نشان‌دهنده‌ی نتایج ضعیف‌تر الگوریتم‌هاست.

بحث

هدف از مطالعه‌ی حاضر، توسعه و ارزیابی مدل‌های ارزیابی و دسته‌بندی پارامتر سطح کنترل آسم با استفاده از ترکیب اطلاعات دموگرافیک بیمار، سوابق پزشکی و متغیرهای بالینی، متغیرهای زیست‌محیطی، متغیرهای عملکرد ریوی و متغیرهای حاصل از پرسش‌نامه‌ها و فرم‌های استاندارد توسعه‌یافته برای کنترل بیماری آسم می‌باشد. به‌همین منظور، با استفاده از انواع روش‌های دسته‌بندی در داده‌کاوی، این پارامتر دسته‌بندی شده و سپس با ارزیابی عملکرد نتایج حاصل از مدل‌ها، الگوریتم xgboost که یک الگوریتم یادگیری گروهی است، به‌عنوان مدل دسته‌بندی با عملکرد بهتر و دقت میانگین ۰/۸۹ و صحت ۰/۷۸ برگزیده شده است. در مقایسه با

مقالات پیشین، Luo و همکاران (۱۳)، به مسئله‌ی پیش‌بینی تضعیف سطح کنترل آسمی با کمک الگوریتم‌های رندم فارست، نزدیک‌ترین همسایگان، ماشین‌بردار پشتیبان و شبکه عصبی پرداختند که در آن مطالعه، الگوریتم نزدیک‌ترین همسایگان با ۰/۷۸ بالاترین دقت و ۰/۵۹/۸ صحت را به‌دست آورده است. Kocsis و همکاران (۲۱) در مطالعه‌ی خود به دسته‌بندی سطح کنترل آسم با سه الگوریتم رندم فارست، بوستینگ و ماشین‌بردار پشتیبان پرداختند که الگوریتم بوستینگ با ۰/۸۴ بهترین دقت را به‌دست آورده است که نتیجه‌ی مشابهی با مطالعه‌ی حاضر داشته که یک الگوریتم بوستینگ به‌عنوان یک الگوریتم یادگیری گروهی نتایج مناسب‌تری نسبت به سایر الگوریتم‌های دسته‌بندی به‌دست آورده است. در این مقاله، سایر شاخص‌های ارزیابی عملکرد گزارش نشده است. همچنین Vliet و همکاران (۶)، در مطالعه‌ی خود برای پیش‌بینی حملات آسمی از الگوریتم نزدیک‌ترین همسایگان استفاده کرده‌اند و دقت مدل ۰/۵۳/۷۷ و صحت آن ۰/۵۸/۵۳ بوده است که نتایج متفاوتی با مطالعه لیو و همکاران (۱۳) داشته که الگوریتم نزدیک‌ترین همسایگان بهترین دقت دسته‌بندی را داشته است. کوپزیک و همکاران نیز (۲۲) برای پیش‌بینی حملات آسمی از الگوریتم رگرسیون لجستیک استفاده نمودند که دقت ۰/۶۵ را به‌دست آورده است که در مقایسه با نتایج الگوریتم رگرسیون در مطالعه‌ی حاضر نتیجه‌ی بهتری به‌دست آورده است. دلایل بهبود نتایج در این مقاله را می‌توان چنین برشمرد:

- مواجهه با چالش پایگاه‌داده‌ای چندمتغیره و چندکلاسه با ویژگی‌های نامتوازن بودن و دارای مقادیر مفقود فراوان
- خوشه‌بندی داده‌ها در مرحله‌ی پیش‌پردازش داده‌ها به منظور افزایش کارایی مدل‌های دسته‌بندی سطح کنترل آسم
- در نظر گرفتن طیف وسیعی از محرک‌های موثر بر سطح کنترل آسم در مدل‌های دسته‌بندی
- نکته‌ی قابل ذکر دیگر این است که پیش‌بینی حملات آسمی یکی از مباحث داغ پژوهشی در مطالعات پیشین می‌باشد (۱۴ و ۶).
- حمله‌ی آسمی واقعه‌ای است که به‌موجب ضعف بلندمدت در سطح کنترل آسم حاصل شده و نشان‌دهنده‌ی تشخیص دیر هنگام ضایعه‌ی

اگر نتایج ارزیابی مدل‌های دسته‌بندی قابل قبول باشد، خروجی آن می‌تواند به یک ابزار الکترونیکی خودمدیریتی آسم متصل شود تا هشدارهای زودهنگام در خصوص سطح کنترل برای بیماران یا پزشکان معالج ارسال شوند. در چنین ابزاری، تمام متغیرهای مستقل باید توسط سیستم خودنظارتی الکترونیکی به‌طور مداوم جمع‌آوری و سپس با کمک مدل دسته‌بندی در مورد سطح کنترل آسم به بیماران آگاهی داده شود. اگر بیمار در سطح کنترلی سبز قرار گیرد، پیشنهاد می‌شود که بیمار برنامه درمانی قبلی را همچنان ادامه دهد. اگر در وضعیت زرد باشد، باید بعد از مشورت با پزشک، برنامه درمانی خود را با تغییر دوز دارو یا تغییر داروهای مصرفی تنظیم مجدد نماید تا به وضعیت مطلوب سبز رنگ ارتقا یابد و اگر در وضعیت قرمز قرار گیرد باید سریعاً اقدامات لازم توسط پزشک معالج انجام شود تا حملات آسمی به بیمار دست ندهد.

نتیجه‌گیری

در این مقاله، پارامتر سطح کنترل آسم به‌عنوان یکی از پارامترهای اثرگذار بر درمان بیماری مزمن آسم، با استفاده از ترکیب اطلاعات دموگرافیک بیمار، سوابق پزشکی، متغیرهای بالینی، فاکتورهای زیست‌محیطی، متغیرهای عملکرد ریوی و متغیرهای حاصل از پرسش‌نامه‌های استاندارد کنترل بیماری آسم دسته‌بندی و سپس با کمک ارزیابی عملکرد نتایج حاصل از مدل‌های مختلف داده‌کاوی، الگوریتم *xgboost* که یک الگوریتم یادگیری گروهی است، بهترین نتیجه را از میان سایر الگوریتم‌ها حاصل نموده است. نتایج تحقیق حاضر و مطالعات پیشین نشان می‌دهد که عملکرد الگوریتم‌های دسته‌بندی سطح کنترل آسم از یک پایگاه داده به یک پایگاه داده دیگر با هم متفاوت است و بستگی به مدل انتخابی پژوهشگران دارد و نمی‌توان مدل واحدی را به عنوان بهترین مدل برگزید. به همین جهت در این مطالعه، نیز طیف متنوعی از الگوریتم‌های دسته‌بندی بر روی پایگاه داده‌ی ایجاد شده، اعمال نموده و با ارزیابی نتایج، بهترین مدل را برگزیدیم. دلیل این رویکرد این است که برای ساخت یک مدل دسته‌بندی قدرتمند برای دسته‌بندی سطح کنترل آسم لازم است

تضعیف کنترل آسم است. در صورتی که، یکی از راه‌های جلوگیری از وقوع حملات آسمی، توسعه‌ی مدل‌های دسته‌بندی سطح کنترل آسم است (۱۳). در واقع با اطلاع از سطح کنترلی که بیمار در آن قرار دارد، می‌توان همواره از وضعیت بیماری اطلاع داشته و برای جلوگیری از تضعیف کنترل آسم از یک مرحله به مرحله‌ی دیگر، اقدام درمانی عاجل و درخور را انجام داد.

نظارت بر کنترل آسم توسط پزشکان در خارج از مراکز بالینی به دلیل نبود ابزارهای موثر، همواره یک چالش است که منجر به درمان‌های واکنشی به‌جای درمان‌های پیشگیرانه از سوی پزشک می‌شود. به‌عبارت دیگر، پزشکان بر روی ویزیت‌های دوره‌ای بیمار تمرکز داشته و به دلیل فقدان ابزارهای مناسب، زمان و منابع کافی و همچنین هزینه‌های گزاف، امکان کنترل بیمار در خارج از مراکز پزشکی و بالینی را ندارند. بیمارستان‌ها نیز اغلب در تلاشند تا تدارکات لازم برای درمان بیماران با آسم حاد را فراهم و روند ترخیص بیماران را تسریع نمایند؛ درحالی که مراقبت‌های مزمن موجب جلوگیری از پذیرش و بستری مجدد بیماران در بیمارستان‌ها می‌شوند. بیماران خانواده‌هایشان نیز در موارد بسیاری نمی‌توانند نشانه‌های تضعیف کنترل آسم را بدون وجود ابزارهای مناسب تشخیص دهند تا زمانی که وضعیت بیمار آنقدر وخیم شود که به مراجعه به بیمارستان نیاز داشته باشد. مفهوم نوین خودمراقبتی مداوم بیماری‌های مزمن بدون وجود ابزارهای کارا که زیربنای آن را مدل‌های دسته‌بندی سطح کنترل آسم تشکیل می‌دهند، برای بیماران و خانواده‌های آنان عملاً امری ناممکن است (۱۳).

مدل پیشنهادی ما در این مقاله، سطح کنترل آسم را با دقت قابل قبولی دسته‌بندی می‌نماید؛ باوجوداین، مداخله‌ای را برای بهره‌برداری از این نتایج در نظر نمی‌گیرد؛ مداخله‌ای که بتواند نتایج حاصل از مدل‌های داده‌کاوی را در قالب ابزارهای فناورانه در اختیار بیماران قرار دهد. نتایج این مدل‌ها می‌توانند به‌صورت یک برنامه سلامت مبتنی بر موبایل در اختیار بیماران قرار گیرند؛ ابزاری که در همه‌ی مکان‌ها و زمان‌ها در اختیار بیمار بوده و مطابق با اهداف خودمدیریتی بیماری‌های مزمنی همچون آسم باشد. به‌عبارت‌دیگر،

اندازه‌گیری پارامترهای آب‌وهوا شناسی و آلودگی هوا

تشکر و قدردانی

این مقاله از رساله دوره دکتری تخصصی دفاع شده در دانشکده مهندسی صنایع و سیستم‌های دانشگاه تربیت مدرس با کد پژوهشی ۲۵۴۹۹۵۴، استخراج شده است. نویسندگان بر خود لازم می‌دانند مراتب تشکر صمیمانه‌ی خود را از کارکنان کلینیک آسم و آلرژی بیمارستان مسیح دانشوری که در انجام و ارتقای کیفی این پژوهش همکاری کرده‌اند، اعلام نمایند.

ترکیبی مناسب از مدل‌ها و ترکیب مناسبی از محرک‌های اثرگذار بر سطح کنترل آسم انتخاب شوند. علاوه بر رویکرد به‌کار گرفته شده در این مقاله، برای مطالعات بعدی اهداف دیگری را می‌توان پیگیری نمود؛ از جمله:

- اضافه‌نمودن سایر متغیرهای تاثیرگذار بر کنترل آسم همانند داده‌های ژنتیکی و نشان‌گرهای زیستی به‌منظور افزایش دقت مدل‌های دسته‌بندی سطح کنترل آسم (به‌ویژه برای دسته‌بندی کلاس قرمز)
- جمع‌آوری داده‌های زیست‌محیطی از محیط داخلی محل زندگی بیمار به‌جای استفاده از داده‌های زیست‌محیطی از ایستگاه‌های

منابع

1. Barnett SB & Nurmagambetov TA. Costs of Asthma in the United States: 2002-2007. *The Journal of Allergy and Clinical Immunology* 2011; 127(1): 145-52.
2. Dexheimer JW, Abramo TJ, Arnold DH, Johnson KB, Shyr Y, Ye F, et al. An Asthma management system in a pediatric emergency department. *International Journal of Medical Informatics* 2013; 82(4): 230-8.
3. Bousquet J, Mantzouranis E, Cruz AA, Ait-Khaled N, Baena-Cagnani CE, Bleecker ER, et al. Uniform definition of Asthma severity, control, and exacerbations: Document presented for the world health organization consultation on severe Asthma. *The Journal of Allergy and Clinical Immunology* 2010; 126(5): 926-38.
4. National Heart, Lung, and Blood Institute. Expert panel report 3: Guidelines for the diagnosis and management of Asthma. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK7232/>. 2007.
5. Toti G, Vilalta R, Lindner P, Lefer B, Macias C & Price D. Analysis of correlation between pediatric Asthma exacerbation and exposure to pollutant mixtures with association rule mining. *Artificial Intelligence in Medicine* 2016; 74(1): 44-52.
6. Vliet DV, Alonso A, Rijkers G, Heynens J, Rosias P, Muris J, et al. Prediction of Asthma exacerbations in children by innovative exhaled inflammatory markers: Results of a longitudinal study. *PLoS One* 2015; 10(3): e0119434.
7. Bateman ED, Buhl R, O'Byrne PM, Humbert M, Reddel HK, Sears MR, et al. Development and validation of a novel risk score for Asthma exacerbations: The risk score for exacerbations. *The Journal of Allergy and Clinical Immunology* 2015; 135(6): 1457-64.
8. Finkelstein J & Wood J. Predicting Asthma exacerbations using artificial intelligence. *Studies in Health Technology and Informatics* 2013; 190(1): 56-8.
9. Farion K, Michalowski W, Wilk S, O'Sullivan D & Matwin S. A tree-based decision model to support prediction of the severity of Asthma exacerbations in children. *Journal of Medical Systems* 2010; 34(4): 551-62.
10. Lee CH, Chen JC & Tseng VS. A novel data mining mechanism considering bio-signal and environmental data with applications on Asthma monitoring. *Computer Methods and Programs in Biomedicine* 2011; 101(1): 44-61.
11. Xu M, Tantisira KG, Wu A, Litonjua AA, Chu JH, Himes BE, et al. Genome Wide Association Study to predict severe Asthma exacerbations in children using random forests classifiers. *BMC Medical Genetics* 2011; 12(90): 1-8.

12. Khasha R, Sepehri MM, Mahdavi SA & Khatibi T. Mobile GIS-based monitoring Asthma attacks based on environmental factors. *Journal of Cleaner Production* 2018; 179(1): 417-28.
13. Luo G, Stone BL, Fassl B, Maloney CG, Gesteland PH, Yerram SR, et al. Predicting Asthma control deterioration in children. Available at: <https://bmcmedinformdecismak.biomedcentral.com/track/pdf/10.1186/s12911-015-0208-9>. 2015.
14. Finkelstein J & Jeong IC. Machine learning approaches to personalize early prediction of Asthma exacerbations. *Annals of the New York Academy of Sciences* 2017; 1387(1): 153-65.
15. Haselkorn T, Zeiger RS, Chipps BE, Mink DR, Szeffler SJ, Simons FE, et al. Recent Asthma exacerbations predict future exacerbations in children with severe or difficult-to-treat Asthma. *The Journal of Allergy and Clinical Immunology* 2009; 124(5): 921-7.
16. Farion KJ, Wilk S, Michalowski W, O'Sullivan D & Sayyad-Shirabad J. Comparing predictions made by a prediction model, clinical score, and physicians: Pediatric Asthma exacerbations in the emergency department. *Applied Clinical Informatics* 2013; 4(3): 376-91.
17. Ko FW, Hui DS, Leung TF, Chu HY, Wong GW, Tung AH, et al. Evaluation of the Asthma control test: A reliable determinant of disease stability and a predictor of future exacerbations. *Respirology* 2012; 17(2): 370-8.
18. Aguinis H, Gottfredson RK & Joo H. Best-Practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods* 2013; 16(2): 270-301.
19. Liu Y & De A. Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International Journal of Statistics in Medical Research* 2015; 4(3): 287-95.
20. Dubey R, Zhou J, Wang Y, Thompson PM & Ye J. Analysis of sampling techniques for imbalanced data: An n = 648 ADNI study. *Neuroimage* 2014; 87(1): 220-41.
21. Kocsis O, Arvanitis G, Lalos A, Moustakas K, Sont JK, Honkoop PJ, et al. Assessing machine learning algorithms for self-management of Asthma, Romania: E-Health and Bioengineering Conference (EHB), 2017.
22. Kupczyk M, Haque S, Sterk PJ, Nizankowska-Mogilnicka E, Papi A, Bel EH, et al. Detection of exacerbations in Asthma based on electronic diary data: Results from the 1-year prospective BIOAIR study. *Thorax* 2013; 68(7): 611-8.

Asthma Control Level Assessment by Moving from the Current Reactive Care Models into a Preventive Approach Based on Fuzzy Clustering and Classification Algorithms

Roghaye khasha¹ (Ph.D.) - Mohammad Mahdi Sepehri² (Ph.D.) - Nasrin Taherkhani³ (Ph.D.)

1 Ph.D. in Information Technology Engineering, Center of Excellence in Healthcare Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran

2 Professor, Department of Healthcare Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran

3 Instructor, Department of Computer Engineering, Faculty of Engineering, Payam-e-Noor University, Saveh, Iran

Abstract

Received: Apr 2020

Accepted: Jul 2020

Background and Aim: Asthma is a common and chronic disease of respiratory tracts. The best way to treat Asthma is to control it. Experts of this field suggest the continues monitoring on Asthma symptoms and adjustment of self-care plan with offering the preventive treatment program to have desired control over Asthma. Presenting these plans by the physician is set based on the control level in which the patient is. Therefore, successful recognition and classification of the disease control level can play an important role in presenting the treatment program to the patient and improves the self-care and strengthens the early interventions to alleviate the Asthma symptoms.

Materials and Methods: Based on this objective, we collected the data of 96 Asthma patients within a 9-month period from a specialized hospital for pulmonary diseases in Tehran. Then we classified the Asthma control level by fuzzy clustering and different types of data mining method within a multivariate dataset with the multi-class response variable.

Results: Our best model resulting from the balancing operations and feature selection on data have yielded the accuracy of 88%.

Conclusion: Our proposed model can be applied in electronic Asthma self-care systems to support the decision in real time and personalized warnings on the possible deterioration of Asthma control. Such tools can centralize the Asthma treatment from the current reactive care models into a preventive approach in which the physician's decisions and therapeutic actions are resulting from the personal patterns of chronic Asthma control and prevention of acute Asthma.

Keywords: Asthma Control, Preventive, Clustering, Classification, Self-Care

* Corresponding Author:
Sepehri M M
Email :
mehdi.sepehri@modares.ac.ir