

بررسی وابستگی وقوع بیماری آسم آلرژیک با ویژگی‌های محیطی با استفاده از کاوش قوانین وابستگی مکانی فازی

فرید کریمی‌پور^۱، یوسف کنعانی سادات^{۲*}

^۱ استادیار گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
fkarimipr@ut.ac.ir

^۲ دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
yousefkanani@ut.ac.ir

(تاریخ دریافت مرداد ۱۳۹۳، تاریخ تصویب آبان ۱۳۹۳)

چکیده

آلرژی یکی از بیماری‌هایی است که به دلیل آلوده شدن محیط اطراف زندگی انسان به عوامل محرک آلرژی، در سالیان اخیر گسترش فزاینده‌ای داشته است. بهترین روش برای درمان بیماری آلرژی، شناسایی عامل آلرژی‌زا و سپس، دوری جستن از آن عامل است. در این میان، عوامل ناشناخته بسیاری هستند که می‌توانند موجب ایجاد آلرژی گردند که اغلب این عوامل از محیط زندگی افراد ناشی می‌شوند. این مقاله، به بررسی وابستگی میان وقوع نوع خاصی از بیماری آلرژی، آسم آلرژیک، با پارامترهای محیطی همچون فاصله از پارک، فاصله از خیابان و آلاینده‌های هوا مانند CO، SO₂، NO₂، PM₁₀، PM_{2.5} و O₃ می‌پردازد. برای این منظور، از روش کاوش قوانین وابستگی چند بُعدی استفاده گردید و برای مدیریت عدم قطعیت موجود در داده‌ها، ابعاد به صورت مجموعه‌های فازی تعریف شدند. قوانین کشف شده نشانگر تاثیر مثبت پارامترهای فاصله از پارک، فاصله از خیابان و آلاینده‌های CO، NO₂، PM₁₀ و PM_{2.5} بر وقوع آسم آلرژیک و عدم تأثیرپذیری آن از آلاینده‌های SO₂ و O₃ هستند.

واژگان کلیدی: آسم آلرژیک، داده‌کاوی، کشف قوانین وابستگی مکانی، قوانین وابستگی چند بُعدی فازی

* نویسنده رابط

۱- مقدمه

امروزه با رشد جوامع، بیماری‌ها نیز از حیث تنوع و تعداد افراد درگیر، در حال افزایش هستند. یکی از بیماری‌هایی که گسترش فزاینده‌ای داشته است، بیماری حساسیت یا آلرژی است [۱]. مبتلایان به این بیماری دارای دستگاه ایمنی فوق هوشیار هستند که نسبت به مواد ظاهراً بی ضرر موجود در محل زندگی، واکنشی بیش از حد معمول نشان می‌دهند. آلرژی مشکل بسیار شایعی است و هر سال ۵٪ به مبتلایان به این بیماری افزوده می‌شود [۲، ۳]. در طول ۵۰ ساله گذشته، بیماری آلرژی در کشورهای مدرن و توسعه یافته در بین کودکان افزایش چشمگیری یافته که این امر، ناشی از آلوده شدن محیط اطراف زندگی به عوامل محرک بوده است [۴]. عوامل مختلفی در ایجاد و تشدید این بیماری دخیل هستند که بسته به نوع جغرافیا، شرایط محیطی و سبک زندگی افراد متغیر هستند [۳]. شناسایی عوامل محرک آلرژی و پیشگیری از مواجهه با مواد حساسیت‌زا بهترین راه برای جلوگیری از بروز حساسیت است [۵]. از آنجا که بخش مهمی از این عوامل، با محیط زندگی انسان مرتبط است، کشف عوامل محیطی موثر بر شیوع آلرژی می‌تواند، نقش بسزایی در کاهش اثرات آن داشته باشد. از این رو می‌توان با جمع‌آوری اطلاعات مناسب از محیط زندگی افراد، نقش عوامل مختلف محیطی را در وقوع و تشدید این بیماری سنجید.

در این مقاله با جمع‌آوری موقعیت مکانی بیماران مبتلا به آسم آلرژیک، و داده‌های محیطی زندگی آنها مانند پارک‌ها و فضاهای سبز، خیابان‌های اصلی و داده‌های ایستگاه‌های سنجش آلودگی هوا، به بررسی امکان ارتباط میان ویژگی‌های محیطی محل زندگی فرد با وقوع آسم آلرژیک در شهر تهران پرداخته شده است. برای کشف این وابستگی، از فرایندهای رایج در داده‌کاوی^۱، مخصوصاً الگوریتم‌های کاوش قوانین وابستگی چند بعدی استفاده گردید. با توجه به حجم بالای داده‌های مورد استفاده و همچنین نقش مولفه مکان در داده‌های مورد مطالعه، داده کاوی مکانی می‌تواند به طور موثری برای دستیابی به ارتباط وقوع آلرژی با عوامل محیطی مورد استفاده قرار گیرد. همچنین به منظور مدیریت عدم قطعیت موجود در

داده‌های مکانی روش کاوش قوانین وابستگی چند بعدی فازی استفاده گردید. از آنجا که بهترین راه برای درمان بیماری‌های آلرژی دوری از عامل آلرژی^۲ است، کشف این وابستگی‌ها و قوانین وابستگی میان عوامل محیطی و وقوع و تشدید آسم آلرژیک می‌تواند برای بیماران مفید باشد و در فرایند تصمیم‌گیری کمک کند.

۲- مروری بر تحقیقات پیشین

از تحقیقات صورت گرفته در زمینه استفاده از داده-کاوی در مطالعه بیماری آلرژی، می‌توان به تحقیق انجام شده توسط Fuang Ng و همکاران در سال ۲۰۰۹ [۴] برای پیش‌بینی آلرژی در میان کودکان در کشور تایوان انجام شد اشاره نمود. آنها از داده‌های آلرژی کودکان زیر ۱۲ سال استفاده کرده و مجموعه‌ای از ۳۰ متغیر پیش‌بین شامل فاکتورهای شخصی، نحوه زندگی، ویژگی‌های خانوادگی، فاکتورهای محرک آلرژی و غیره را مد نظر قرار دادند. در این تحقیق از سه مدل پیش‌بینی کننده شامل درخت تصمیم‌گیری، شبکه عصبی و ماشین بردار پشتیبان^۳ استفاده شد. بر اساس نتایج به دست آمده، مدل درخت تصمیم‌گیری با دقت طبقه‌بندی ۷۰٪ پیش‌بینی بهتری را ارائه داد. همچنین در سال ۲۰۱۲، Youssef Agha و همکاران [۶] از قابلیت‌های داده‌کاوی برای پیش‌بینی بروز آلرژی در بین دانش‌آموزان مدرسه ابتدایی بهره بردند. هدف از این تحقیق پی بردن به رابطه میان وقوع آلرژی در میان بچه‌های مدرسه ابتدایی و مشاهدات روزانه هواشناسی بالای سطح زمین (دما، رطوبت نسبی، نقطه شبنم) و آلودگی هوای روزانه (ریزگردها، سولفور دی اکسید، نیتروژن دی اکسید، کربن مونو اکسید و اوزن) با استفاده از رگرسیون منطقی بود. با وجود نتایج رضایت‌بخش هر دوی این تحقیقات، هیچ یک از آنها، جنبه‌های مکانی داده‌ها در وقوع آلرژی را در نظر نگرفته‌اند.

در این مقاله با در نظر گرفتن جنبه مکانی داده‌ها، از روش کشف قوانین وابستگی برای بررسی وابستگی وقوع بیماری آسم آلرژیک با آلاینده‌های هوا و دیگر عوامل محیطی دیگر، استفاده شده است؛ همچنین برای مدیریت

^۲ Allergen

^۳ Support Vector Machine

^۱ Data Mining

شوند. شناسایی الگوهای تکراری نقش مهمی در شناسایی وابستگی‌ها، همبستگی‌ها و بسیاری دیگر از ارتباط‌های جذاب میان داده‌ها دارند. کاوش مجموعه اقلام تکراری، به کشف وابستگی‌ها و همبستگی‌های بین اقلام در پایگاه داده‌های بزرگ تراکنشی یا رابطه‌ای می‌انجامد [۱۲]. قوانین وابستگی ابزاری برای نمایاندن وابستگی بین داده‌های ثبت شده در پایگاه داده را فراهم می‌کنند که باعث تسهیل فرایند تصمیم‌گیری می‌شوند؛ و به دلیل حجم وسیعی از اطلاعات که امروزه در بخش‌های مختلف ذخیره می‌شود، این موضوع بسیار جذاب شده است [۱۱-۱۳].

اگر مجموعه $I = \{I_1, I_2, \dots, I_m\}$ مجموعه‌ای از اقلام باشد و T مجموعه‌ای از تراکنش‌ها باشد به طوری که هر تراکنش زیرمجموعه‌ای از I است، یک قانون وابستگی به صورت $A \Rightarrow B [s, c]$ تعریف می‌شود، که در آن، مجموعه‌های A و B زیرمجموعه‌های غیرتهی از I هستند و اشتراکی با هم ندارند. پشتیبان s درصدی از تراکنش‌ها است که هر دو مجموعه A و B را شامل می‌شوند و اطمینان c بیانگر درصدی از تراکنش‌های شامل A است که مجموعه B را نیز شامل می‌شوند [۸، ۱۲].

Agrawal و همکاران نخستین الگوریتم کاوش قوانین وابستگی را در سال ۱۹۹۳ برای کشف الگوها در پایگاه داده تراکنشی ارائه کردند [۱۴]. از آنجا که ایده مورد استفاده آنها، به دنبال کشف وابستگی بین اقلام خریداری شده توسط مشتریان از فروشگاه بود، "تحلیل سبد خرید" نیز نامگذاری شد. قانون "مشتری که قلم A را می‌خرد با احتمال ۹۰٪ قلم B را نیز خریداری خواهد کرد"، مثالی از این قوانین است.

در پایگاه داده تراکنشی $D = \{T_1, T_2, \dots, T_n\}$ که T_i نشان دهنده یک تراکنش در پایگاه داده است و در مجموعه اقلام $I = \{I_1, I_2, \dots, I_m\}$ که I_j نشان دهنده یک نوع از اقلام است، هر تراکنش T شامل زیرمجموعه‌ای از اقلام می‌شود ($T \subseteq I$). اگر A و B مجموعه‌ای غیرتهی از اقلام در تراکنش T باشند که هیچ اشتراکی با هم ندارند ($A \cap B = \phi, A \subseteq I, B \subseteq I$)، یک قانون وابستگی در پایگاه داده D به صورت $A \Rightarrow B$ تعریف می‌-

عدم قطعیت موجود در داده‌ها از مجموعه‌های فازی بهره گرفته شده است.

کاوش قوانین وابستگی یکی از رایج‌ترین مدل‌های مطالعه شده در داده‌کاوی است. در سال‌های اخیر، کشف قوانین وابستگی از پایگاه داده به یکی از موضوعات مهم و فعال تحقیقاتی در زمینه داده‌کاوی تبدیل شده است [۷، ۸]. از تحقیقات انجام شده در زمینه کشف قوانین وابستگی می‌توان به تحقیق Wang در سال ۲۰۱۰ [۸] برای کشف قوانین وابستگی چند بعدی اشاره کرد. وی الگوریتمی بر مبنای الگوریتم اپریوری^۱ ارائه کرد که از آن برای تجزیه و تحلیل وابستگی اجناس فروخته شده استفاده کرد. در این مقاله ثابت شده است که با اعمال این روش به محیط تجارت واقعی، میزان فروش به طور چشمگیری افزایش پیدا می‌کند. Mennis و همکاران در سال ۲۰۰۵ [۹] به کاوش قوانین وابستگی در داده‌های مکانی زمانمند پرداختند. در این تحقیق، کاوش قوانین وابستگی برای کشف روابط مکانی و زمانی میان مجموعه‌ای از متغیرهای توصیف کننده شرایط اجتماعی-اقتصادی و تغییرات پوشش گیاهی در منطقه Colorado آمریکا از سال ۱۹۷۰ تا سال ۱۹۹۰ استفاده شده است. Shua و همکاران در سال ۲۰۰۸ [۱۰] به کشف قوانین وابستگی در داده‌های مکانی زمانمند پرداختند. آنها وابستگی میان داده‌های پوشش گیاهی و تغییرات آب و هوا در شمال شرقی چین را مورد مطالعه قرار دادند. برای این منظور، پس از پیش‌پردازش‌های اعمال شده بر روی داده‌ها، از الگوریتم اپریوری برای تولید قوانین وابستگی استفاده نمودند. در سال ۲۰۰۳، Ladner و همکاران [۱۱] وابستگی مکانی میان داده‌هایی مانند نوع خاک، روابط جهتی یا هندسی و غیره را بررسی کردند. در این تحقیق تکنیک‌های داده کاوی مکانی و داده کاوی فازی برای مقابله با عدم قطعیت موجود در داده‌های مکانی با هم ترکیب و گسترش داده شدند.

۳- قوانین وابستگی

الگوهایی که به صورت مکرر در مجموعه داده‌ها تکرار می‌گردند (مانند خرید همزمان نان و شیر در مجموعه داده‌های خرید از فروشگاه) الگوهای تکراری نامیده می‌-

^۲ Support
^۳ Confidence

^۱ Apriori

معیارهای مختلفی برای اندازه‌گیری همبستگی بین A و B وجود دارد. یکی از معیارهای اندازه‌گیری همبستگی، معیار کولژینسکی است. این روش که در سال ۱۹۲۷ توسط Kulczynski پیشنهاد شد حاصل میانگین دو احتمال شرطی احتمال وقوع A به شرط B و احتمال وقوع B به شرط A است [۱۸].

$$Kulc(A, B) = \frac{1}{2}(P(A|B) + P(B|A)) \quad (3)$$

مقدار این معیار در بازه [0,1] است و هر چقدر مقدار معیار بزرگتر و به ۱ نزدیکتر باشد ارتباط نزدیکتری بین A و B وجود دارد.

۳-۱- itemset ها در داده‌های چند بعدی

در یک پایگاه داده رابطه‌ای که شامل m سطر و n ستون است، (d_1, d_2, \dots, d_n) ابعاد مجموعه داده‌ها هستند و d_j نشان‌دهنده j-امین بعد مجموعه داده‌ها است. $(v_{11}, v_{12}, \dots, v_{1n})$ نشان‌دهنده مقادیر ابعاد مختلف برای i-امین سطر از مجموعه داده‌ها است که مقدار j-امین بعد در i-امین سطر را نشان می‌دهد.

در مجموعه داده‌های چندبعدی، هر قلم به صورت زیر تعریف می‌شود [۱۶]:

$$I_{ij} = (d_{ij}, v_{ij}), \quad 1 \leq j \leq n, 1 \leq i \leq m \quad (4)$$

و یک itemset مجموعه‌ای است که به وسیله تعدادی قلم ساخته می‌شود.

۳-۲- قوانین وابستگی چندبعدی

قانونی مانند قانون "نان، خرید) => (شیر، خرید)" را قانون وابستگی یک‌بعدی و یا قانون وابستگی درون بعدی^۵ می‌نامند، به دلیل اینکه تنها دارای یک بعد یا گزاره^۶ "خرید" است که چند بار در قانون تکرار شده است. اگر قوانین وابستگی دو یا بیش از دو بعد داشته باشند، در این صورت با یک قانون وابستگی چندبعدی روبرو هستیم. برای مثال، قانون وابستگی "ماشین، خرید) => (بالا ، درآمد)^(جوان ، سن)" که دارای ۳ بعد "خرید"، "درآمد"

شود. A و B به ترتیب مقدم^۱ و نتیجه^۲ (تالی) قانون وابستگی هستند. معمولاً میزان اعتمادپذیر بودن، درستی و جذابیت قوانین وابستگی کشف شده با دو معیار پشتیبان و اطمینان سنجیده می‌شود که طبق روابط زیر به دست می‌آیند.

$$sup(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$conf(A \Rightarrow B) = P(B|A) = \frac{P(A \cup B)}{P(A)} \quad (2)$$

مقدار پشتیبان حاکی از اهمیت بیرونی^۳ یا فرکانس رخداد یک قانون وابستگی است و مقدار اطمینان حاکی از صحت ذاتی^۴ یا عقلانی بودن یک قانون وابستگی است [۱۰]. با تعریف دو حد آستانه حداقل پشتیبان و حداقل اطمینان، قوانینی که این حد آستانه‌ها را پشتیبانی می‌کنند و به قوانین قوی موسوم هستند، شناسایی می‌شوند. شناسایی قوانین وابستگی قوی در دو مرحله انجام می‌گیرد [۷، ۸، ۱۲، ۱۵، ۱۶]:

۱- شناسایی همه مجموعه اقلام تکراری در پایگاه

داده مورد نظر: در این مرحله مجموعه اقلام تکراری که حداقل پشتیبان را دارند شناسایی می‌شوند.

۲- تولید قوانین وابستگی قوی از مجموعه اقلام

تکراری شناسایی شده: در این مرحله قوانینی که حداقل اطمینان تعریف شده را دارند انتخاب می‌شوند.

بخش اول که هسته الگوریتم نیز هست، نقش مهمی

در کارایی داده‌کاوی دارد.

معیارهای اطمینان و پشتیبان برای حذف قوانین وابستگی ضعیف کافی نیستند. فاکتور همبستگی معیار دیگری برای ارزیابی همبستگی بین A و B است. به این ترتیب می‌توان قوانین همبسته‌ای به صورت زیر شناسایی کرد [۱۲، ۱۷]:

$$A \Rightarrow B \text{ [support, confidence, correlation]}$$

- ۱ Premise
- ۲ Consequence
- ۳ Extrinsic Importance
- ۴ Intrinsic Correctness

^۵ Intradimensional Association Rule
^۶ Predicate

• برخلاف قوانین وابستگی غیر مکانی که وابستگی‌ها به طور صریح در پایگاه داده وجود دارند، کاوش قوانین وابستگی مکانی به دنبال پیدا کردن الگویی در میان روابط مکانی است که این روابط به طور صریح در پایگاه داده نیستند و در یک چارچوب مکانی زمین مرجع قرار دارند، بنابراین روابط مکانی باید قبل از کاوش قوانین وابستگی از داده‌ها استخراج شوند [۹]، [۱۰، ۲۱].

• گزاره‌های مکانی اغلب شامل داده‌های عددی هستند (مانند فاصله متریک)، درحالی‌که روش‌های مرسوم کاوش قوانین وابستگی برای داده‌های طبقه‌بندی‌شده^۳ طراحی شده‌اند و برای داده‌های عددی مناسب نیستند. این مشکل را می‌توان با طبقه‌بندی داده‌های عددی به دسته‌های ترتیبی و سپس کاوش این داده‌های ترتیبی برای کشف قوانین وابستگی حل کرد [۲۲، ۲۳]. برای مثال فاصله متریک را می‌توان به دو طبقه "نزدیک" و "دور" تقسیم کرد.

۳-۵- کاوش قوانین وابستگی مکانی فازی

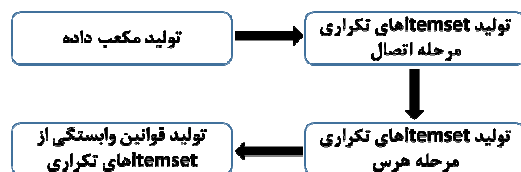
نیاز برای مدیریت اطلاعات نادقیق^۴ مربوط به اطلاعات مکانی امری اجتناب‌ناپذیر است [۲۴]. ارزش GIS به عنوان ابزار تصمیم‌گیری، به توانایی تصمیم‌گیرنده‌ها برای ارزیابی قابلیت اعتماد اطلاعاتی که تصمیم‌شان بر مبنای آن اطلاعات است، بستگی دارد. بنابراین کاربران GIS باید قادر به ارزیابی طبیعت و درجه خطا در پایگاه داده‌های مکانی، دنبال کردن این خطا از طریق عملگرهای مکانی و برآورد صحت محصول خروجی باشند [۱۱].

بسیاری از عملیات‌ها روی داده‌های مکانی با این فرض که عوارض، ویژگی‌ها و روابط بین آنها از پیش به صورت دقیق مشخص شده است، انجام می‌گیرد. با این حال عدم دقت اغلب در موقعیت عوارض و در نسبت‌دهی مقادیر ویژگی‌ها به عوارض وجود دارد و ممکن است در مراحل مختلف تدوین و گردآوری داده ایجاد شود [۱۱]، همچنین اگر داده‌ای که شامل محدوده‌ای از مقادیر است، وجود

و "سن" است یک قانون وابستگی چندبعدی است. قوانین وابستگی چندبعدی که بعد تکراری ندارند، مانند قانون بالا را قوانین وابستگی برون بعدی^۱ می‌نامند. درنهایت، قوانین وابستگی مانند قانون "پرینتر، خرید" => (لپ‌تاپ، خرید)∩(جوان، سن)" که دارای بعد تکراری باشند، قوانین وابستگی دارای ابعاد ترکیبی^۲ گفته می‌شوند [۸، ۱۲، ۱۵].

۳-۳- کاوش قوانین وابستگی چندبعدی

برای تعیین وابستگی میان داده‌های چندبعدی، در مرحله اول باید در مورد ابعادی که مقادیر عددی دارند گسسته‌سازی صورت گیرد و مقادیر عددی به دسته‌های گسسته تقسیم شده و تبدیل به مقادیر کیفی شوند. سپس تولید itemsetهای تکراری صورت می‌گیرد. این مرحله با فرایندی که در الگوریتم اپریوری کلاسیک [۱۹] وجود دارد متفاوت است. در مرحله اتصال، مجموعه کاندیداها (C_k) با اتصال L_{k-1} هایی که $k-2$ قلم یکسانی دارند، به وجود می‌آید. در صورتی که بعد از اتصال، کاندیدایی دارای اقلامی با بعد یکسان باشد، آن کاندیدا حذف خواهد شد. مرحله هرس نیز مشابه الگوریتم اپریوری کلاسیک انجام می‌گیرد؛ به این صورت که هر کاندیدا که دارای زیرمجموعه غیرتکراری باشد، یعنی یک زیرمجموعه ($k-1$) عضوی از یک کاندیدای C_k عضو L_{k-1} نباشد، آن کاندیدا حذف می‌گردد [۸] (شکل ۱).



شکل ۱- روند کاوش قوانین وابستگی چندبعدی

۳-۴- کاوش قوانین وابستگی مکانی

قوانین وابستگی که یکی از گزاره‌ها در مقدم یا تالی شامل یک رابطه مکانی باشد، قوانین وابستگی مکانی نامیده می‌شوند [۹، ۲۰]. دو مشکل مهم در کاوش قوانین وابستگی در داده‌های مکانی وجود دارد:

^۳ Categorical
^۴ Uncertain

^۱ Interdimensional Association Rule
^۲ Hybrid-dimensional Association Rule

اگر A و B مجموعه‌های فازی باشند و d یک قلم داده باشد، مقادیر پشتیبان، اطمینان و همبستگی در کاوش قوانین وابستگی فازی از روابط زیر به دست می‌آیند [۱۵]:

$$\text{support}(A) = \frac{\sum_{i=1}^r \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}}{r} \quad (6)$$

$$\text{support}(A \Rightarrow B) = \text{support}(A \cup B) = \frac{\sum_{i=1}^r \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{r} \quad (7)$$

$$\text{confidence}(A \Rightarrow B) = \frac{\sum_{i=1}^r \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{\sum_{i=1}^r \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}} \quad (8)$$

این فرمول‌ها برای محاسبه میزان پشتیبان، اطمینان و همبستگی قوانین استفاده می‌شوند که در آن‌ها، μ_x مقدار عضویت x و r تعداد داده‌های موجود در پایگاه داده است. واضح است که قانون وابستگی عمومی در واقع حالت خاصی از قانون وابستگی فازی است که مقدار عضویت هر عنصر در این حالت برابر ۰ یا ۱ است [۱۰].

۴- پیاده‌سازی و روند انجام تحقیق

این مقاله به بررسی وابستگی میان وقوع و تشدید آسم آلرژیک با ویژگی‌های محیطی با کشف قوانین وابستگی مکانی فازی می‌پردازد. منطقه مورد مطالعه شهر تهران بوده و داده‌هایی مانند غلظت آلاینده‌های هوا، نقشه پارک-ها و نقشه خیابان‌های اصلی جمع‌آوری گردید تا به بررسی تاثیر این عوامل بر روی بیماری آسم آلرژیک بپردازیم (شکل ۲).



شکل ۲- نقشه خیابان‌های اصلی، پارک‌ها و ایستگاه‌های سنجش آلودگی هوای تهران

معماری کلی تحقیق در شکل ۳ نشان داده شده است.

داشته باشد مشکل مرز تیز ایجاد می‌شود که باعث عدم قطعیت در مرزها می‌شود [۲۵].

تعدادی از محققان GIS و پایگاه داده‌های مکانی، مدل‌سازی عدم قطعیت موجود در داده‌های مکانی را با استفاده از روش مجموعه‌های فازی بررسی کرده‌اند [۲۶-۲۸]. مدل‌سازی اشیاء جغرافیایی با مرزهای نامشخص مثالی از کاربرد مجموعه‌های فازی است [۲۹]. اولین رویکرد سازگار در استفاده از تئوری مجموعه‌های فازی که در GIS می‌تواند استفاده شود، توسط Robinson توسعه داده شد [۳۰-۳۲].

در کاوش قوانین وابستگی مکانی، عدم قطعیت در هر مرحله از فرایند وجود دارد: از پیش‌پردازش داده‌ها و مفهومی‌سازی داده‌ها تا استخراج قوانین وابستگی [۱۰]. استفاده از مجموعه‌های فازی در کشف قوانین وابستگی از پایگاه داده مکانی به علت توانایی مجموعه‌های فازی در مدل کردن عدم قطعیتی که در داده‌ها وجود دارد، مفید بوده [۳۳] و راه حلی برای مشکل مرز تیز در داده‌ها می‌باشد [۲۵].

۳-۶- قوانین وابستگی فازی

برای کاوش و کشف قوانین وابستگی بامعنا تر و قابل اعتمادتر، ابعاد مجموعه داده مورد بررسی باید به صورت مجموعه‌های فازی تعریف شوند [۱۵، ۱۷].

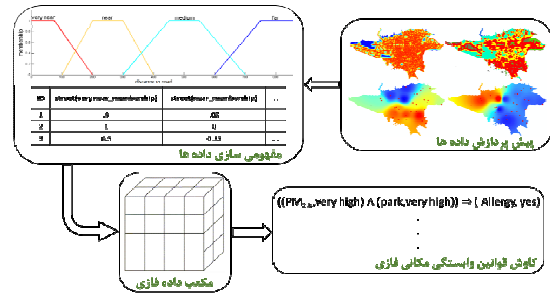
مجموعه فازی با برجسب فازی بامعنا نمایش داده می‌شود و تابع عضویت برای نسبت‌دهی مقدار عضویت به هر عضو مجموعه استفاده می‌شود. برای مثال تابع عضویت برای مجموعه فازی "خیلی نزدیک" می‌تواند به صورت زیر برای فاصله تعریف شود:

$$\text{very_near}(x) = \begin{cases} 1 & x \leq 75 \\ \frac{200-x}{125} & 75 \leq x \leq 200 \\ 0 & x \geq 200 \end{cases} \quad (5)$$

یک قانون وابستگی فازی از ابعاد فازی تشکیل می‌شود [۱۷]. به عنوان مثال یک قانون وابستگی چند بعدی فازی با گزاره‌های distance_to_NO_2 و disease به صورت زیر تعریف می‌شود:

$$\text{Distance_to (road, very near)} \wedge \text{NO}_2 (X, \text{high}) \Rightarrow \text{disease} (X, \text{allergy})$$

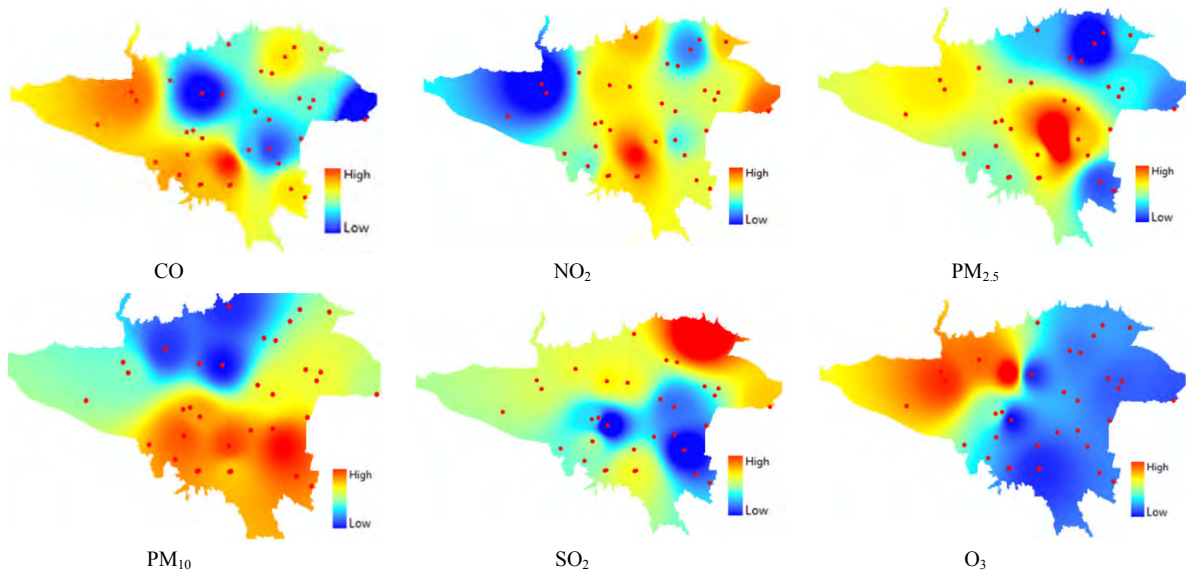
زمانی آذر ماه سال ۱۳۹۲ هستند و به صورت ساعتی اندازه گیری شده‌اند. برای این مشاهدات، داده‌های از دست رفته پر می‌شوند، و داده‌های دارای نویز فیلتر می‌شوند. برای کاهش حجم زیاد داده‌ها به پارامتر آلودگی ماهانه، برای هر آلاینده، ماکزیمم غلظت آلودگی برای یک روز به عنوان آلودگی روزانه در نظر گرفته شده و با میانگین‌گیری از غلظت‌های روزانه برای یک ماه، آلودگی متوسط ماهانه برای هر ایستگاه محاسبه گردید. این مقادیر برای تولید نقشه آلودگی ماهانه برای هر آلاینده، از طریق تابع درونیابی مکانی Kriging [۳۴] در محیط ArcGIS، استفاده شدند. شکل ۴ نقشه آلودگی برای آلاینده‌های مختلف را در آذر ماه نشان می‌دهد.



شکل ۳- معماری کلی تحقیق

۴-۱- پیش‌پردازش داده‌ها

داده‌های آلودگی هوا مربوط به ایستگاه‌های سنجش آلودگی هوای شهرداری تهران و سازمان محیط زیست از این سازمان‌ها اخذ گردید. این داده‌ها شامل مقدار غلظت آلاینده‌های CO ، NO_2 ، SO_2 ، O_3 ، $PM_{2.5}$ و PM_{10} در بازه



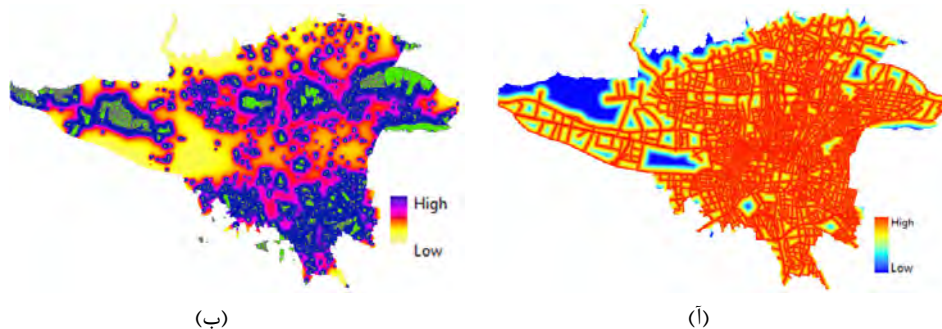
شکل ۴- نقشه آلودگی آلاینده‌های مختلف در شهر تهران (آذر ۱۳۹۲)

(شکل ۵.ب). برای تهیه این لایه، تاثیر پارک‌ها برای هر نقطه با رابطه زیر محاسبه شد:

$$T_j = \sum \frac{A_i}{d_{ij}^2} \quad (9)$$

که در این رابطه T_j تاثیر پارک‌ها بر نقطه d_{ij} فاصله نقطه j از پارک i و A_i مساحت پارک i است.

همچنین با وارد کردن نقشه خیابان‌های اصلی در ArcGIS، لایه‌ای تهیه شد که در این لایه فاصله هر نقطه، از نزدیکترین خیابان تا آن نقطه مشخص است (شکل ۵.آ). برای مدلسازی تاثیر پارک‌ها بر نقاط مختلف لایه‌ای تهیه شد، به این صورت که هر چه یک نقطه به پارک نزدیک‌تر باشد تاثیرپذیری نقطه از پارک بیشتر می‌شود و هر چه مساحت پارک بیشتر باشد، اثرگذاری پارک بیشتر است



شکل ۵- نقشه‌های تولید شده برای (آ) فاصله از خیابان و (ب) تاثیر پارکها



شکل ۶- لایه نقطه‌ای موقعیت مکانی جمعیت مورد مطالعه

در نهایت با بررسی پرونده پزشکی ۱۰۰۰ نفر از مراجعه-کنندگان به "مرکز طبی کودکان تهران"، در ماه آذر، موقعیت مکانی محل زندگی و داشتن یا نداشتن آسم آلرژیک برای هر فرد مشخص شد. علاوه بر این، موقعیت مکانی افراد، به صورت یک لایه اطلاعاتی نقطه‌ای، تهیه گردید (شکل ۶) و با برهم‌نهی این لایه با لایه‌های تولید شده در اشکال ۴ و ۵، مقدار مربوط به هر لایه برای هر فرد مشخص و در جدول اطلاعاتی بیماران ذخیره گردید. جدول ۱ قسمتی از جدول استخراج شده است.

جدول ۱- قسمتی از جدول اطلاعاتی استخراج شده برای بیماران

شماره	تاثیر پارک	فاصله از خیابان	CO	NO ₂	PM _{2.5}	SO ₂	PM ₁₀	O ₃	آسم آلرژیک
۱	۳,۳۵	۱۰۱,۹۸	۵۹,۰۸	۸۱,۵۹	۲۰۱,۴۵	۶۱,۳	۱۱۳,۲۸	۲۷,۸۶	ندارد
۲	۲,۹۵	۴۸۰,۱	۸۴,۸۸	۵۳,۴	۱۵۷,۲۴	۱۱۴,۳۷	۱۰۲	۲۹,۳۷	ندارد
۳	۲,۸۲	۱۸۰,۲۷	۸۴,۹۸	۵۳,۰۵	۱۵۷,۰۳	۱۱۳,۸۴	۹۹,۷۹	۲۹	ندارد
۴	۳,۲۹	۹۰	۷۳,۴۶	۱۰۱,۹	۲۰۷,۱۶	۵۹,۳۱	۱۱۹,۰۴	۲۶,۶۹	دارد
۵	۲,۵	۲۷۸,۹۲	۶۶,۹۴	۱۰۳,۱۱	۱۷۶,۳۸	۸۷,۲۸	۷۵,۵۶	۴۴,۷۵	ندارد
۶	۲,۲۲	۴۰۶,۰۷	۹۱,۸۱	۳۸,۷۲	۱۹۹,۱۲	۸۴,۵۵	۹۸,۸۱	۷۸,۵۴	ندارد
۷	۵,۴۷	۷۰	۹۰,۷۱	۷۴,۴۶	۱۸۵,۲۶	۷۶,۳۴	۱۲۵,۸۴	۳۶,۷۹	دارد
۸	۲,۴۴	۵۲۷,۷۳	۸۱,۷۳	۶۰,۲۶	۱۹۷,۰۸	۸۰,۲۸	۹۸,۶۵	۶۶,۰۹	ندارد
۹	۲,۰۵	۲۸۳,۰۱	۸۱,۹۲	۶۲,۷۱	۱۹۶,۳۳	۷۸,۸۳	۱۰۱,۲۳	۶۳,۰۶	ندارد
۱۰	۲,۳۶	۲۰۶,۱۵	۷۵,۴۵	۸۱,۱۳	۱۹۳,۶۴	۷۱,۴۶	۱۱۲,۹۶	۴۶,۶۲	دارد

۴-۲- مفهومی سازی داده‌ها

بین میزان آلودگی آلاینده‌های مختلف داشت، از شاخص کیفیت هوا (AQI) استفاده می‌کنیم. به طور کلی AQI شاخصی جهت گزارش کیفیت هوا است. این شاخص مردم را از کیفیت هوا (پاک بودن یا آلوده بودن آن) آگاه می‌سازد و اثرات سلامتی مرتبط با آن را ارائه می‌کند. به عبارت دیگر AQI به اثرات سلامتی ناشی از مواجهه با هوای آلوده می‌پردازد. به منظور درک آسان، شاخص کیفیت هوا به شش دسته طبقه‌بندی شده است که هر دسته را به سطوح مختلف سلامت انسان مربوط می‌سازد؛ این شش دسته در جدول ۲ نشان داده شده‌اند. برای محاسبه AQI از رابطه (۱۰) استفاده می‌شود که

همانطور که در بخش ۳ گفته شد، ورودی الگوریتم کاوش قوانین وابستگی باید مقادیر طبقه‌بندی شده باشد. بنابراین داده‌های مربوط به جدول اطلاعاتی بیماران باید طبقه‌بندی شوند. علاوه بر این، برای مدیریت عدم قطعیت موجود در داده‌ها و مشکل مرز تیز در طبقه‌بندی داده‌ها، دسته‌ها به صورت مجموعه فازی تعریف شدند. با توجه به این که میزان غلظت مجاز آلاینده‌های مختلف برای سلامتی انسان متفاوت است، برای این که بتوان مقایسه‌ای

C_p = غلظت اندازه گیری شده برای آلاینده p است.
 BP_{Hi} = نقطه شکستی که بزرگتر یا مساوی C_p است.
 BP_{Lo} = نقطه شکستی که کوچکتر یا مساوی C_p است.
 I_{Hi} = مقدار AQI منطبق با BP_{Hi}
 I_{Lo} = مقدار AQI منطبق با BP_{Lo}

پارامترهای مورد استفاده در این رابطه از جدول ۲ که نقاط شکست برای AQI را نشان می دهند به دست می آید:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + I_{Lo} \quad (10)$$

در این رابطه:

I_p = شاخص کیفیت هوا (AQI) برای آلاینده p است.

جدول ۲- نقاط شکست برای AQI [۳۵]

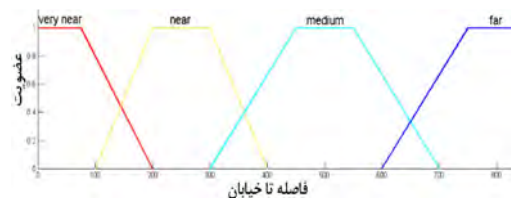
نقاط شکست						AQI	طبقه بندی کیفیت هوا
NO2 (ppb)	SO2 (ppb)	CO (ppm)	PM _{2.5} (µg/m ³)	PM ₁₀ (µg/m ³)	O3 (ppm)		
۰-۵۳	۰-۳۵	۰-۴,۴	۰-۱۵,۴	۰-۵۴	۰-۰,۰۵۹	۰-۵۰	خوب
۵۴-۱۰۰	۳۶-۷۵	۴,۵-۹,۴	۱۵,۵-۴۰,۴	۵۵-۱۵۴	۰,۰۶-۰,۰۷۵	۵۱-۱۰۰	متوسط
۱۰۱-۳۶۰	۷۶-۱۸۵	۹,۵-۱۲,۴	۴۰,۵-۶۵,۴	۱۵۵-۲۵۴	۰,۰۷۶-۰,۰۹۵	۱۰۱-۱۵۰	ناسالم برای گروه های حساس
۳۶۱-۶۴۹	۱۸۶-۳۰۴	۱۲,۵-۱۵,۴	۶۵,۵-۱۵۰,۴	۲۵۵-۳۵۴	۰,۰۹۶-۰,۱۱۵	۱۵۱-۲۰۰	ناسالم
۶۵۰-۱۲۴۹	۳۰۵-۶۰۴	۱۵,۵-۳۰,۴	۱۵۰,۵-۲۵۰,۴	۳۵۵-۴۲۴	۰,۱۱۶-۰,۳۷۴	۲۰۱-۳۰۰	خیلی ناسالم
۱۲۵۰-۲۰۴۹	۶۰۵-۱۰۰۴	۳۰,۵-۵۰,۴	۲۵۰,۵-۵۰۰	۴۲۵-۶۰۴	-	۳۰۱-۵۰۰	خطرناک

توابع عضویت روی مقادیر جدول اطلاعاتی بیماران اعمال می شود و مقادیر عددی، به وسیله برچسب های فازی توصیف می شوند. برای مثال مقدار عضویت ویژگی "فاصله از خیابان" برای هر فرد در هر دسته، در جدول ۳ نشان داده شده است.

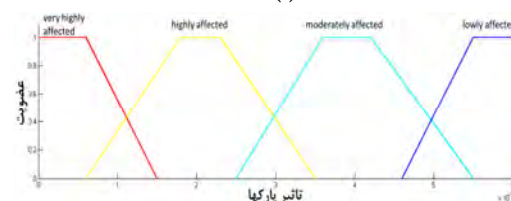
جدول ۳- مقدار عضویت فاصله از خیابان برای هر فرد در دسته های مختلف

شماره	فاصله از خیابان	very near	near	medium	far
۱	۱۰۱,۹۸	۰,۷۸	۰,۰۲	۰	۰
۲	۴۸۰,۱	۰	۰	۱	۰
۳	۱۸۰,۲۷	۰,۱۶	۰,۸	۰	۰
۴	۹۰	۰,۸۸	۰	۰	۰
۵	۲۷۸,۹۲	۰	۱	۰	۰
۶	۴۰۶,۰۷	۰	۰	۰,۷	۰
۷	۷۰	۱	۰	۰	۰
۸	۵۲۷,۷۳	۰	۰	۱	۰
۹	۲۸۳,۰۱	۰	۱	۰	۰
۱۰	۲۰۶,۱۵	۰	۱	۰	۰

ما دسته های بالا را در دسته های فازی "low"، "moderate"، "high" و "very high" ادغام کردیم. همچنین تابع عضویت نشان داده شده در شکل ۷.۱ برای طبقه بندی "فاصله از خیابان" به دسته های فازی "far"، "medium"، "near" و "very near" استفاده شده است. تابع عضویت شکل ۷.۲ نیز برای طبقه بندی "تاثیر پارک" به دسته های فازی "lowly affected"، "moderately affected"، "highly affected" و "very highly affected" استفاده شد.



(آ)



(ب)

شکل ۷- توابع عضویت برای طبقه بندی (آ) فاصله از خیابان و (ب) تاثیر پارکها

۴-۳- ساخت مکعب داده فازی

اتخاذ ساختار مکعب داده چند بعدی که یک سازمان - دهی رایج داده برای داده کاوی در ساختارهای انباره داده است، برای کاوش کارآمدتر قوانین وابستگی چند بعدی، سودمند است [۱۱, ۱۲]. یک مکعب داده n بعدی یک پایگاه داده n بعدی است که هر بعد یک خصیصه^۱ را نشان می‌دهد و هر سلول مکعب داده، تعداد سطریهایی که با مقادیر خصیصه n بعدی مربوطه مطابقت دارد را ثبت می‌کند که می‌توان از آن‌ها برای محاسبه پشتیبان و اطمینان قوانین وابستگی چند بعدی استفاده کرد [۸, ۱۲]. استفاده از مکعب داده به صورت فازی روش بهینه‌ای را برای کشف دانش فراهم کرده است [۱۱]. در این تحقیق برای استفاده از مزایای مکعب داده در حالت فازی، در هر سلول، مجموع مینیمم مقدار تابع عضویت داده‌ها در دسته‌های مربوط محاسبه شده و در سلول مربوطه قرار می‌گیرد. برای مثال جدول ۴ مقادیر عضویت پنج قلم اطلاعاتی ۱ تا ۵ را در دسته‌های فازی a_1 و a_2 از ویژگی a و در دسته‌های فازی b_1 و b_2 از ویژگی b نشان می‌دهد. عنصر (a_1, b_1) از مکعب داده در جدول ۵ محاسبه شده است. برای همه ترکیبات دسته‌های a و b این مقادیر محاسبه می‌شود و مکعب داده فازی ساخته می‌شود. جدول ۶ مکعب داده فازی ساخته شده برای داده‌های جدول ۴ را نشان می‌دهد.

جدول ۶- مکعب داده ۲ بعدی ساخته شده برای داده‌های جدول ۴

	a2	a1	
جمع	۲,۸۱	۱,۵	b1
	۳,۳۱	۲,۰	b2
	۶,۱۲	۳,۶	جمع

به این طریق یک مکعب داده فازی دارای ۹ بعد برای داده‌های این تحقیق ساخته شد به صورتی که ابعاد آن ویژگی‌های محیطی (فاصله از خیابان، تاثیر پارک‌ها و آلاینده‌های هوا) و وضعیت آلرژي برای فرد است.

۴-۴- کاوش قوانین وابستگی مکانی فازی

پس از ساخت مکعب داده فازی، با استفاده از الگوریتم کاوش قوانین وابستگی چند بعدی فازی، وابستگی میان آسم آلرژیک و ویژگی‌های محیطی کشف می‌شوند. در این تحقیق حد آستانه حداقل پشتیبان و حداقل اطمینان به ترتیب ۰.۵ و ۰.۴۰ تعریف شدند. از آنجایی که ما به دنبال قوانینی هستیم که طرف دوم قانون، بیانگر وجود آلرژي باشد، ما این گونه قوانین را حفظ کرده و از بقیه صرف نظر می‌کنیم، مانند قانون:

(allergy, yes) \rightarrow (PM_{2.5}, very high)

۵- نتایج تحقیق

با انجام رویه توضیح داده شده در بخش ۴ بر روی داده‌ها، تعداد ۶۰ قانون وابستگی کشف شد که تعدادی از این قوانین در جدول ۷ نشان داده شده است. به عنوان مثال قانون ۵ با میزان پشتیبان ۰.۶۵۵ و میزان اطمینان ۰.۷۵۵۲ بیانگر این است که ۶.۵۵ درصد از کل جمعیت مورد مطالعه، در مکان‌هایی سکونت داشته‌اند که میزان آلاینده NO₂ و PM_{2.5} خیلی زیاد بوده و به علت فاصله کم تا پارک‌های اطراف محل زندگی، تاثیرپذیری خیلی زیادی از پارک داشته‌اند و به بیماری آسم آلرژیک مبتلا شده‌اند و ۷۵.۵۲ درصد از جمعیتی که در منطقه با میزان آلاینده NO₂ و PM_{2.5} خیلی زیاد و تاثیرپذیری خیلی زیاد از پارک، سکونت داشته‌اند، آسم آلرژیک گرفته‌اند.

جدول ۴- مقادیر عضویت در خصیصه‌های a و b

سطر	$\mu_{a_1}(a)$	$\mu_{a_2}(a)$	$\mu_{b_1}(b)$	$\mu_{b_2}(b)$
۱	۱	۰	۰	۱
۲	۰.۶۶	۰.۲۷	۱	۰
۳	۰.۵۳	۰.۴۱	۰.۸۳	۰.۱۱
۴	۰.۸۷	۰.۱	۰	۱
۵	۰.۳۳	۰.۶۱	۰.۸۳	۰.۱۱

جدول ۵- محاسبه عنصر (a_1, b_1) برای مکعب داده

سطر	$\mu_{a_1}(a)$	$\mu_{b_1}(b)$	مینیمم
۱	۱	۰	۰
۲	۰.۶۶	۱	۰.۶۶
۳	۰.۵۳	۰.۸۳	۰.۵۳
۴	۰.۸۷	۰	۰
۵	۰.۳۳	۰.۸۳	۰.۳۳
جمع			۱.۵۲

^۱ Attribute

جدول ۷- قسمتی از قوانین کشف شده

شماره	قوانین وابستگی	پشتیبان	اطمینان	همبستگی
۱	$[(PM_{2.5}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۱۴,۶۱	۴۳,۵۲	۰,۶۳
۲	$[(PM_{2.5}, \text{very high}), (\text{park_efct}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۹,۳۵	۵۵,۴۳	۰,۶۱
۳	$[(PM_{2.5}, \text{very high}), (PM_{10}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۹,۴۷	۵۱,۶۲	۰,۶
۴	$[(PM_{2.5}, \text{very high}), (PM_{10}, \text{very high}), (\text{park_efct}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۶,۱۵	۷۲,۶	۰,۶۲
۵	$[(PM_{2.5}, \text{very high}), (NO_2, \text{very high}), (\text{park_efct}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۶,۵۵	۷۵,۵۲	۰,۶۴
۶	$[(NO_2, \text{very high}), (CO, \text{very high}), (\text{park_efct}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۵,۳۲	۸۱,۶۳	۰,۶۴
۷	$[(PM_{10}, \text{very high}), (NO_2, \text{very high}), (CO, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۷,۲۵	۶۸,۵۵	۰,۶۲
۸	$[(PM_{2.5}, \text{very high}), (PM_{10}, \text{very high}), (NO_2, \text{very high}), (\text{park_efct}, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۵,۸۴	۷۸,۶۴	۰,۶۴
۹	$[(PM_{2.5}, \text{very high}), (PM_{10}, \text{very high}), (NO_2, \text{very high}), (CO, \text{very high})] \rightarrow (\text{allergy}, \text{yes})$	۶,۶۹	۷۱,۳۳	۰,۶۲
۱۰	$[(PM_{10}, \text{very high}), (NO_2, \text{very high}), (CO, \text{very high}), (\text{road}, \text{very near})] \rightarrow (\text{allergy}, \text{yes})$	۵,۳۹	۷۱,۸۴	۰,۵۹
۱۱	$[(PM_{2.5}, \text{very high}), (PM_{10}, \text{very high}), (NO_2, \text{very high}), (CO, \text{very high}), (\text{road}, \text{very near})] \rightarrow (\text{allergy}, \text{yes})$	۵,۳۹	۷۳	۰,۶

بر اساس قوانین کشف شده پارامترهای محیطی مانند نزدیکی به پارک، نزدیکی به خیابان و افزایش آلاینده‌های $PM_{2.5}$ ، PM_{10} ، CO و NO_2 بر روی بروز و تشدید آسم آلرژیک افراد تاثیرگذار هستند، درحالیکه آلاینده‌های SO_2 و O_3 تاثیر چندانی در وقوع آسم آلرژیک ندارند. از طرف دیگر قوانین دارای " $(\text{park_efct}, \text{very high})$ " که حداقل دارای یک گزاره آلاینده خیلی زیاد هستند (مانند قوانین ۲، ۴ و ۶) میزان اطمینان بیشتری نسبت به قوانینی که فقط شامل آلاینده هستند (مانند قوانین ۱، ۳ و ۷)، دارند. تحقیقات صورت گرفته در زمینه آلودگی هوا و آلرژی تایید کننده قوانین کشف شده هستند. بر اساس این تحقیقات آلاینده‌ها علاوه بر این که می‌توانند عامل بروز آسم آلرژیک شوند، همچنین می‌توانند باعث انتقال آسان-تر گرده‌های گیاهان به سیستم تنفسی انسان شوند [۳۶]. این نشان‌دهنده انطباق این قوانین با واقعیت است.

۶- نتیجه‌گیری و پیشنهادات

در این مقاله به بررسی وابستگی میان وقوع آسم آلرژیک با عوامل محیط زندگی پرداخته شد. برای استخراج وابستگی‌های ناشناخته بین پارامترهای محیطی مانند فاصله از خیابان، فاصله از پارک و آلاینده‌های CO

از SO_2 ، NO_2 ، PM_{10} ، $PM_{2.5}$ و O_3 با بیماری آسم آلرژیک، از روش کاوش قوانین وابستگی مکانی فازی استفاده گردید. نتایج به دست آمده بیانگر این واقعیت هستند که در نظر گرفتن موقعیت مکانی بیمار، کشف وابستگی‌های قابل اعتمادتر را ممکن می‌سازد. آلاینده‌های SO_2 و O_3 تاثیر چندانی در وقوع و تشدید آسم آلرژیک ندارند، در حالی که سایر عوامل بررسی شده، باعث تشدید این بیماری می‌شوند. با تحلیل پارامتر اطمینان و همبستگی قوانین به دست آمده مشاهده می‌شود که جاهایی که به پارک نزدیک هستند و همچنین آلودگی خیلی زیادی دارند، مکان‌های مستعدتری برای شیوع بیماری آلرژی هستند. همچنین طبق قوانین کشف شده در بین آلاینده‌ها، قوانینی که شامل تاثیر خیلی زیاد پارک هستند و مقدار آلودگی خیلی زیاد دارند دارای اطمینان و همبستگی بیشتری هستند که نشانگر تاثیر زیاد ذرات آلاینده‌ها در ترکیب با گرده گیاهان بر روی سیستم تنفسی انسان و بروز آسم هستند. با توجه به این که بهترین راه برای درمان بیماری‌های آلرژیک دوری از عامل محرک است، بنابراین برای بهبود آسم آلرژیک و عدم تشدید این بیماری، توصیه می‌شود بیمار از سکونت در نزدیکی پارک‌ها و مکان‌های آلوده اجتناب کنند.

کرد. همچنین به دلیل این‌که گرده افشانی گیاهان و آلودگی هوا در زمان‌های مختلف متفاوت است انجام زمانمند این تحلیل می‌تواند به کشف قوانین قابل اعتمادتری منجر شود.

با انجام این بررسی عوامل محیطی که بر تشدید و بروز بیماری آسم آلرژیک تاثیر گذار هستند کشف شدند. با معلوم شدن عوامل تاثیر گذار بر آسم آلرژیک، از این عوامل می‌توان در مدلسازی مکانی این بیماری استفاده

مراجع

- [1] I. Zöllner, S. Weiland, I. Piechotowski, T. Gabrio, E. Von Mutius, B. Link, et al., "No increase in the prevalence of asthma, allergies, and atopic sensitisation among children in Germany: 1992–2001," *Thorax*, vol. 60, pp. 545-548, 2005.
- [2] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA special publication*, vol. 351, p. 309, 1974.
- [3] M. Asher, U. Keil, H. Anderson, R. Beasley, J. Crane, F. Martinez, et al., "International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods," *European respiratory journal*, vol. 8, pp. 483-491, 1995.
- [4] H.-F. Ng, H. Fathoni, and I.-C. Chen, "Prediction of Allergy Symptoms among Children in Taiwan Using Data Mining," 2009.
- [5] J .A. Douglass and R. E. O Hehir, "Diagnosis, treatment and prevention of allergic disease: the basics," *Medical journal of Australia*, vol. 185, p. 228, 2006.
- [6] A. YoussefAgha, W. Jayawardene, D. Lohrmann, and G. El Afandi, "Application of Data Mining Techniques to Predict Allergy Outbreaks among Elementary School Children," 2012.
- [7] H. Sug, "Discovery of multidimensional association rules focusing on instances in specific class," *International Journal of mathematics and Computers in Simulation*, vol. 5 ,pp. 250-257, 2011.
- [8] F. Wang, "Application of multidimensional association rule techniques in manufacturing resource planning system," in *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2010 Seventh International Conference on, 2010, pp. 1433-1437.
- [9] J .Mennis and J. Liu, "Mining association rules in spatio-temporal data," in *Proceedings of the 7th International Conference on GeoComputation*, 2003.
- [10] H. Shua, X. Zhub, and S. Daic, "Mining association rules in geographical spatio-temporal data," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B2. Beijing 2008, 2008.
- [11] R. Ladner, F. E. Petry, and M. A. Cobb, "Fuzzy set approaches to spatial data mining of association rules," *Transactions in GIS*, vol. 7, pp. 123-138, 2003.
- [12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publishers Inc., 2011.
- [13] R. Rastogi and K. Shim, "Mining optimized association rules with categorical and numeric attributes," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 14, pp. 29-50, 2002.
- [14] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD Record*, 1993, pp. 207-216.
- [15] R. Intan, "A Proposal of Fuzzy Multidimensional Association Rules," *Jurnal Informatika*, vol. 7, pp. pp. 85-90, 2007.

- [16] W.-X. Xu and R.-J. Wang, "A fast algorithm of mining multidimensional association rules frequently," in *Machine Learning and Cybernetics, 2006 International Conference on*, 2006, pp. 1199-1203.
- [17] R. Intan, A. Handojo, and O. Yenty Yuliana, "Mining Fuzzy Multidimensional Association Rules using Fuzzy Decision Tree Induction Approach," *International Journal of Computer and Network Security (IJCNS)*, vol. 1, 2009.
- [18] S. Kulczynski, "Die Pflanzenassoziationen der Pieninen," *Bulletin International de l'Academie Polonaise des Sciences et des Lettres, Classe des Sciences Mathematiques et Naturelles B*, pp. 57-203., 1927.
- [19] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases, VLDB, 1994*, pp. 487-499.
- [20] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," presented at the 4th International Symposium on Large Spatial Databases, Berlin, 1995.
- [21] S. Shekhar and S. Chawla, *Spatial databases: a tour vol. 2003: prentice hall Upper Saddle River, NJ, 2003*.
- [22] G. Piatetsky-Shapiro, "Discovery, analysis and presentation of strong rules," *Knowledge discovery in databases*, pp. 229-238, 1991.
- [23] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in *ACM SIGMOD Record*, 1996, pp. 1-12.
- [24] M. Goodchild and A. Gopal, *The Accuracy of Spatial Databases*. London: Taylor and Francis, 1990.
- [25] S. Jain, A. P. S. Jain, and A. Jain, "An Assessment of Fuzzy Temporal Association Rule Mining," *International Journal*, 2013.
- [26] S. Gale, "Inexactness, Fuzzy Sets, and the Foundations of Behavioral Geography ",**Geographical Analysis*, vol. 4, pp. 337-349, 1972.
- [27] Y. Leung, "Locational choice: A fuzzy set approach," *Geographical Bulletin*, vol. 15, pp. 28-34, 1979.
- [28] J. S. Pipkin, "Fuzzy sets and spatial choice," *Annals of the Association of American Geographers*, vol. 68, pp. 196-204, 1978.
- [29] P. A. Burrough and A. Frank, *Geographic objects with indeterminate boundaries vol. 2: CRC Press*, 1996.
- [30] V. B. Robinson, "Some implications of fuzzy set theory applied to geographic databases," *Computers, Environment and Urban Systems*, vol. 12, pp. 89-97, 1988.
- [31] V. B. Robinson, "Interactive machine acquisition of a fuzzy spatial relation," *Computers & Geosciences*, vol. 16, pp. 857-872, 1990.
- [32] V. B. Robinson and A. U. Frank, "About different kinds of uncertainty in collections of spatial data," in *Auto-Carto 7, American Society for Phtogrammetry and Remote Sensing and American Congress on Surveying and Mapping*, 1985.
- [33] S. U. Calargun and A. Yazici, "Fuzzy Association Rule Mining from Spatio-temporal Data," in *Computational Science and Its Applications-ICCSA 2008*, ed: Springer, 2008, pp. 631-646.
- [34] H. Wackernagel, *Multivariate geostatistics: Springer*, 2003.
- [35] D. Mintz, *Technical Assistance Document for the Reporting of Daily Air Quality-the Air Quality Index (AQI): US Environmental Protection Agency, Office of Air Quality Planning and Standards*, 2012.
- [36] J. Bartra, J. Mullol, A. Del Cuillo, I. Dávila, M. Ferrer, I. Jáuregui, et al., "Air pollution and allergens," *J Investig Allergol Clin Immunol*, vol. 17, pp. 3-8, 2007.