Original Article

Depiction of Health
2020; 11(1): 20-30
http://dohweb.tbzmed.ac.ir

Detection of the Association Rules of the Occurrence of Brucellosis in Humans Using Spatial Data Mining

Mohammad Abbasi*[1], Farid Karimipour[2], Sima Gholipour[1]

Article Info:

Article History:
Received: 05/25/2019
Accepted: 10/12/2019
Published: 03/20/2020

Keywords:
Brucellosis
Data Mining
Association Rule
Geographical Information
Knowledge

Abstract

Background and Objectives: Brucellosis is the most common disease in humans and animals, where its distribution is roughly global, so the control of this disease has an important impact on human health and livestock industry. The most important causes of this disease are environmental parameters (location). In this regard, identifying and discovering the relationship between environmental parameters and the outbreak and incidence of the disease can be effective in controlling and providing preventive measures and thus reducing the cost of treatment for humans and animals.

Material and Methods: In this research, using spatial data and multidimensional association rules, the association between the occurrence of Brucellosis with environmental and human parameters such as temperature, humidity, altitude, land use, occupation and history of contact with the livestock and use of non-pasteurized dairy has been investigated with emphasis on location. The statistical population of this study is people with brucellosis in Hamedan province in 1394.

Results: The results of this study indicate that factors of location and non location included temperature, altitude, humidity, history used of non-pasteurized dairy, history of contact with infected animals sick people, job and land use, can be effective in the spread of the disease.

Conclusion: The results of the study show that the factors of temperature and elevation have had more effects in the discovered laws, so that areas with higher temperatures and lower altitudes than other areas include more patients. Also, people with agricultural and livestock jobs occupy the highest rates of patients in the laws that were discovered.

Citation: Abbasi M, Karimipour F, Gholipour S. Detection of the Association Rules of the Occurrence of Brucellosis in Humans Using Spatial Data Mining. Depiction of Health 2020; 11(1): 20-30.

1. University College of Omran-Toseeh, Hamedan, Iran (**Email**: mohammadabbasi@ut.ac.ir)
2. Faculty of Surveying and Geospatial Engineering, University of Tehran, Iran

TUOMS PRESS
Tabriz University of Medical Sciences

© 2020 The Author(s). This work is published by Depiction of Health as an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

مقاله پژوهشی

کاوش قوانین وابستگی وقوع تب مالت در انسان با استفاده از داده‌کاوی ویژگی‌های مکانی

محمد عباسی[1*]، فرید کریمی‌پور[2]، سیما قلی‌پور[1]

چکیده

زمینه و اهداف: تب مالت شایع‌ترین بیماری مشترک انسان و دام است که توزیع مکانی آن تقریبا جهانی است، لذا کنترل این بیماری تاثیر مهمی در سطح جهانی بر روی سلامت انسان و صنعت دامداری دارد. از عوامل مهم شیوع این بیماری پارامترهای محیطی (مکان) می‌باشد. در این راستا شناسایی و کشف روابط پارامترهای محیطی با نحوه شیوع و وقوع بیماری می‌تواند در کنترل و ارائه راه‌کارهای پیشگیرانه و در نتیجه کاهش هزینه‌های درمان انسان و دام موثر باشد.

مواد و روش‌ها: در این تحقیق، با استفاده از داده‌کاوی مکانی و قوانین وابستگی چندبعدی، ارتباط بین وقوع بیماری تب مالت در انسان با پارامترهای محیطی و انسانی مانند دما، رطوبت، ارتفاع، کاربری زمین، شغل افراد و سابقه تماس آنها با دام و سابقه استفاده از مواد لبنی غیرپاستوریزه با تاکید برمکان بررسی شده است. جامعه آماری این تحقیق افراد مبتلا به تب مالت در سال ۱۳۹۴ در استان همدان می‌باشد.

یافته‌ها: نتایج حاصل از این مطالعه بیانگر این است که عوامل مکانی و غیرمکانی از جمله دما، ارتفاع منطقه، رطوبت، سابقه استفاده از لبنیات غیرپاستوریزه، سابقه تماس افراد بیمار با دام‌های آلوده، شغل افراد مبتلا شده و پوشش منطقه‌ای که افراد مبتلا شده در آن زندگی می‌کنند، می‌توانند در شیوع بیماری تب‌مالت موثر باشند.

نتیجه‌گیری: نتایج پژوهش نشان می‌دهد که عوامل دما و ارتفاع دارای تاثیرات بیشتری در قوانین کشف شده بوده‌اند، به گونه‌ای که مناطق دارای دمای زیادتر و ارتفاع پایین‌تر نسبت به دیگر مناطق، شامل افراد بیمار بیشتری بوده‌اند. همچنین افراد دارای شغل کشاورزی و دامداری در قوانین کشف شده بیشترین آمار مبتلایان را به خود اختصاص داده‌اند.

کلیدواژه‌ها: تب‌مالت، داده‌کاوی، قاعده، وابستگی، اطلاعات جغرافیایی، دانش

نحوه استناد به این مقاله: عباسی م، کریمی‌پور ف، قلی‌پور س. کاوش قوانین وابستگی وقوع تب مالت در انسان با استفاده از داده‌کاوی ویژگی‌های مکانی. تصویر سلامت ۱۳۹۹؛ ۱۱(۱): ۲۰-۳۰.

۱. موسسه آموزش عالی عمران و توسعه، همدان، ایران (**Email**: mohammadabbasi@ut.ac.ir)

۲. دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران، ایران

حقوق برای مؤلف(ان) محفوظ است. این مقاله با دسترسی آزاد در تصویر سلامت تحت مجوز کریتیو کامنز (http://creativecommons.org/licenses/bync/4.0/) منتشر شده که طبق مفاد آن هرگونه استفاده غیر تجاری تنها در صورتی مجاز است که به اثر اصلی به نحو مقتضی استناد و ارجاع داده شده باشد.

مقدمه

تب مالت انسانی شایع‌ترین نوع بیماری‌های مشترک انسان و دام در جهان است. به جز تعداد معدودی از کشورها که توانسته‌اند ارگانیسم بیماری را از منابع حیوانی ریشه‌کن نمایند، وقوع بیماری در اغلب نقاط دنیا دیده می‌شود. انتقال بیماری معمولا از طریق تماس مستقیم با حیوانات آلوده و یا مصرف محصولات آلوده مانند شیر و پنیر تازه صورت می‌گیرد (۱-۴).

تب‌مالت به علت عوارض طولانی مدتی که دارد، به بیماری هزار چهره معروف شده است. این وضعیت باعث شده است که پزشکان در تشخیص صحیح و به موقع آن، دچار مشکل شوند و بیماران نیز دچار ناتوانایی‌های جسمی و روحی مزمن و عودکننده گردند. از یک طرف هزینه‌های تشخیص و درمان این بیماری بار مالی زیادی را بر نهادهای دولتی و درمانی کشور تحمیل کرده و از طرف دیگر ضرر و زیان اقتصادی ناشی از معدوم کردن دام‌های آلوده نیز از پی‌آمدهای نامطلوب این بیماری محسوب می‌شود (۵).

طبق گزارش وزارت بهداشت در سال ۱۳۹۱، استان‌های آذربایجان شرقی، همدان، لرستان، مرکزی، خراسان جنوبی، آذربایجان غربی و کرمانشاه دارای آلودگی بسیار بالا (۳۱-۴۱ در هر صد هزار نفر) و استان‌های خراسان رضوی، کردستان و زنجان دارای آلودگی بالا (۳۰-۲۱ در هر صد هزار نفر) بوده‌اند. این بیماری در هر دو جنس مذکر و مونث دیده می‌شود. بروز بیماری در جنس مذکر کمی بیشتر از جنس مونث بوده است. شغل مبتلایان یک عامل موثر در وقوع بیماری بوده و در خانم-های خانه‌دار، دامداران و کشاورزان بیشتر دیده شده است. از نظر زمانی هم این بیماری بیشتر در فصول بهار و تابستان هم-زمان با فصل زایش و شیردهی دام‌ها مشاهده شده است (۶).

در این راستا با توجه به حجم بالای داده‌های ذخیره شده و لزوم بررسی و تفسیر آنها جهت کشف دانش، استفاده از داده-کاوی ضروری است. داده‌کاوی هر نوع استخراج دانش و یا الگو از داده‌های موجود در پایگاه داده است که این دانش‌ها و الگوها ضمنی و مستتر در داده‌ها هستند. به صورت تخصصی داده‌کاوی سلامت، دانشی میان رشته‌ای، جدید و در حال رشد است که با تلفیق علوم مختلف مانند پزشکی، آمار، کامپیوتر، یادگیری ماشین، فناوری‌های اطلاعات مکانی و غیره دانش نهفته در داده‌ها را کشف نموده و به توسعه تحقیقات پزشکی و تصمیم‌گیری‌های بالینی کمک می‌نماید (۷-۱۰).

با توجه به اینکه امروزه مکان، عاملی کلیدی در تفسیر پدیده‌های مختلف می‌باشد، بحث داده‌کاوی مکانی مطرح است. بررسی کشف دانش در داده‌کاوی مکانی به ویژگی‌های انحصاری که داده‌کاوی مکانی را از داده‌کاوی سنتی مجزا می‌کند، تمرکز دارد. فعالیت‌های عمده و نیازهای تحقیقاتی در مطالعات داده‌کاوی مکانی بسیار مهم بوده و این امر، منجر به ایجاد روش‌ها و برنامه‌های کاربردی، موضوعات و چالش‌های داده‌کاوی مکانی می‌گردد، که نشان‌دهنده این است که داده‌کاوی مکانی یک میدان امیدوارکننده با نتایج تحقیقات پربار و بسیاری از مسائل چالش برانگیز است (۱۱).

در زمینه استفاده از داده‌کاوی در مطالعه بیماری تب‌مالت، محققین به تجزیه و تحلیل توزیع زمانی و مکانی تب‌مالت انسانی در آذربایجان در سال‌های ۱۹۹۵ تا ۲۰۰۹ با استفاده از آمار فضایی و زمانی-مکانی پرداختند. در این مطالعه تجزیه و تحلیل خوشه‌ای زمانی و مکانی با استفاده از شاخص Moran's I محلی و تست Ederer-Myer-Mantel محاسبه شده است (۱۲).

در تحقیقی، محققین با تحلیل و بررسی بیماری تب‌مالت و شناسایی نقش برخی پارامترهای اقلیمی موثر در بروز و اپیدمی آن، به تهیه نقشه‌های پهنه‌بندی تب‌مالت انسانی و دامی و نهایتا به تعیین کانون‌های شیوع آن پرداخته‌اند. در این پژوهش بین دو گروه داده آمار افراد و دام‌های مبتلا و داده‌های آب و هوایی اعم از دما، بارش و رطوبت ضریب همبستگی پیرسون محاسبه شده و سپس با استفاده از تکنیک زمین‌آمار پهنه‌بندی به روش IDW در نرم‌افزار GIS، نقشه‌های جغرافیایی مربوطه ترسیم شده که نشان داده است عامل دما با وقوع تب‌مالت ارتباط مستقیم داشته و نسبت به سایر مولفه‌های اقلیمی، موثرترین عامل بر بروز و شیوع این بیماری است (۱۳).

در این پژوهش با استفاده از موقعیت مکانی افراد بیمار و دام های مبتلا، به تفکیک روستا و محل زندگی بیماران و همچنین داده های هواشناسی، مدل رقومی زمین (DEM)، کاربری زمین و شغل افراد مبتلا شده، به بررسی امکان ارتباط میان ویژگی های محیطی محل زندگی افراد با وقوع بیماری تب مالت در استان همدان پرداخته‌شده است. با توجه به حجم بالای داده ها و همچنین نقش مولفه مکان در داده های مورد مطالعه، داده‌کاوی مکانی می‌تواند به طور موثری برای دستیابی به ارتباط وقوع تب مالت با عوامل محیطی مورد استفاده قرار گیرد. برای کشف وابستگی ها، از فرآیند های رایج در داده‌کاوی، به خصوص الگوریتم‌های کاوش قوانین وابستگی چندبعدی استفاده شده است.

مواد و روش‌ها

در این تحقیق، با توجه به اینکه استان همدان یکی از استان-های دارای آمار بالای بیماری تب مالت در کشور می‌باشد، (۶) به عنوان منطقه مورد مطالعه انتخاب شده است.

در این پژوهش، برای بررسی وابستگی میان وقوع تب‌مالت در انسان با ویژگی‌های محیطی، مجموعه داده‌ها شامل مکان زندگی افراد مبتلا به بیماری تب مالت در استان همدان، نقشه دمای هوا، میزان رطوبت و ارتفاع متوسط هر منطقه، شغل و سابقه تماس با دام، جمع‌آوری و نقشه‌های مربوط به هر داده

تولید شدند تا به بررسی تاثیر این عوامل بر روی بیماری تب‌مالت با استفاده از داده‌کاوی مکانی پرداخته شود.

در این تحقیق بطورکلی داده‌ها به دو دسته کلی طبقه‌بندی می‌شوند که شامل داده‌های مربوط به افراد مبتلا به بیماری تب-مالت (موقعیت افراد بیمار، سابقه تماس با دام، سابقه استفاده از لبنیات غیرپاستوریزه و شغل) و داده‌های مکانی از شرایط محیطی (دما، رطوبت، ارتفاع و کاربری زمین) در منطقه مورد مطالعه می‌باشند.

در این پژوهش داده‌های مربوط به افراد مبتلا به بیماری تب‌مالت در سال ۱۳۹۴ از معاونت بهداشتی استان همدان اخذ گردید که این داده‌ها شامل، آدرس، اطلاعات سن، جنسیت، شغل و سابقه تماس با دام می‌باشند. همچنین داده‌های مربوط به دمای هوا و رطوبت میانگین در ۸ ایستگاه، در منطقه مطالعاتی، از سازمان هواشناسی استان همدان اخذ گردید.

در این پژوهش برای تولید نقشه‌های مورد نیاز در محیط نرم‌افزار ArcGIS از روش‌های تراکم کرنل و روش درونیابی کریجینگ و جهت کشف دانش از روش داده‌کاوی قوانین وابستگی در محیط نرم‌افزار SPSS استفاده شده است که بطور خلاصه شرح داده شده‌اند.

روش تراکم کرنل دارای این مزیت است که با نمایش سطحی می‌تواند مدل پیوسته‌تری را از الگوهای نقاط کانونی ارائه دهد. روش برآورد تراکم کرنل که در سال ۱۹۵۶ توسط رزنبلات ارائه گردید، توجه ویژه‌ای را در برآورد غیرپارامتریک تراکم احتمالی به خود جلب نموده‌است. برآورد تراکم کرنل روش پرطرفداری جهت تحلیل داده‌های یک بعدی و دو بعدی می‌باشد که اطلاعات مفیدی را در مورد عوارض در داده‌ها ارائه می‌دهد. این روش قادر است به سرعت و به صورت بصری نقاط کانون را از پایگاه داده بزرگ شناسایی کرده و در نتیجه خروجی آماری و رضایت بخشی را ارائه دهد.

برآورد $p(x)$ از تراکم در X با استفاده از رابطه زیر محاسبه گردد:

(۱)

$$R_i = p(x) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(X - X_i)$$

که در آن K_σ تابع کرنل است (۱۴).

کریجینگ یک روش برآورد زمین آماری است که بر پایه میانگین متحرک وزن‌دار استوار است. به‌طوریکه می‌توان گفت که این روش بهترین برآوردکننده خطی ناریب می-باشد. ویژگی کریجینگ در آن است که در عین ناریب بودن، واریانس تخمین نیز در کم‌ترین مقدار می‌باشد. بطورکلی کریجینگ دارای دو وظیفه اصلی است:

واریوگرافی: تعیین کمیت ساختار مکانی داده‌ها (برازش مدل وابسته به مکان روی داده‌هاست)

تولید پیش‌بین: ساخت یک پیش‌بین برای مقادیر مجهول در یک موقعیت خاص

فرمول مورد استفاده در این روش:

$$Z(S_o) = \sum \lambda_i \times Z(S_i) \qquad (۱)$$

که در آن: $Z(S_o)$ موقعیت نقاط مجهول، λ_i (فاصله و خودهمبستگی) مقدار اندازه‌گیری شده در نقطه iام و $Z(S_i)$ مقادیر اندازه‌گیری شده در نقطه iام می‌باشد (۱۵).

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 شده‌است (۱۶).

منظور از قوانین وابستگی، کشف وابستگی بین عناصری است که رخداد آنها در یک زمان است یا به عبارت دیگر تکراری است. برای مثال، اجناسی که در یک فروشگاه احتمال خرید آنها با هم در یک تراکنش خرید زیاد است. این اقلام دارای وابستگی هستند که این وابستگی‌ها به صورت $A \Longrightarrow B$ نمایش داده می‌شود. به A مقدم و به B موخر یا نتیجه گفته می‌شود. کشف مجموعه عناصر تکرار شونده، به کشف وابستگی بین عناصر در مجموعه داده-ها با حجم زیاد منجر می‌شود. از ارزیاب‌های مهم در قوانین وابستگی که به منظور ارزیابی صحت قوانین کشف شده مورد استفاده قرار می‌گیرد، اطمینان (Confidence) و پشتیبان (Support) هستند.

اطمینان: زمانی که وقوع یک پدیده به وقوع پدیده‌های دیگری منجر می‌شود، احتمال وقوع با استفاده از این معیار اندازه‌گیری می‌شود.

پشتیبان: اگر وقوع دو پدیده با هم اتفاق بیفتد، میزان احتمال وقوع آن با این معیار، اندازه‌گیری و میزان درصد وقوع آنها با هم، با عدد پشتیبان نشان داده می‌شود.

قوانین وابستگی دارای یک آستانه حداقل پشتیبان و یک آستانه حداقل اطمینان هستند که با توجه به این مقدار آستانه، معناداری قوانین تشخیص داده می‌شود (۱۷).

در این تحقیق از الگوریتم‌های پیشینی (Apriori) و کاوش قوانین وابستگی چندبعدی استفاده شده است.

الگوریتم پیشینی یک روش تکرارپذیر است که k-itemsetها برای یافتن (K+1)-itemset ها را مورد استفاده قرار می‌دهد و از دو بخش اتصال (join) و هرس (prune) تشکیل شده است. زمانی که itemset های تکرارپذیر از بین تراکنش‌ها در بانک اطلاعات به دست

آمدنـد، استخراج قـوانین وابـستگی محکم از آنها امکان‌پذیر است که با استفاده از رابطه زیر انجام می‌شود:

$$Confidence(A \Longrightarrow B) = P(B|A) = support_count(A \cup B)/support_count(A)$$

براساس رابطه ۳، قوانین وابستگی می‌تواند به صورت زیر تولید شود:

برای هر itemset تکرارپذیر l، همه زیر مجموعه‌های غیر تهی آن ایجاد شود.

برای هر زیر مجموعه غیرتهی S از l قانون:

$$S \Rightarrow (l - s): if\ (support_count(l)/ suport_count(s)) \geq min_conf$$

Where min_conf is the minimum confidence threshold.

یعنی اگر تعـداد تکرارهـای Itemset انتخـابی l بـر تعـداد تکرارهـای زیرمجموعـه انتخـاب شـده از آن، از مقـدار min_conf بزرگتـر باشـد. آنگـاه s⇒ (l-s) یک قانون وابستگی است. چون قـوانین از itemsetهـای تکرارپـذیر تولیـد مـی‌شـوند، هـر کـدام بـه صورت خودکار، حداقل پشتیبان مورد نظر را دارند (۱۸).

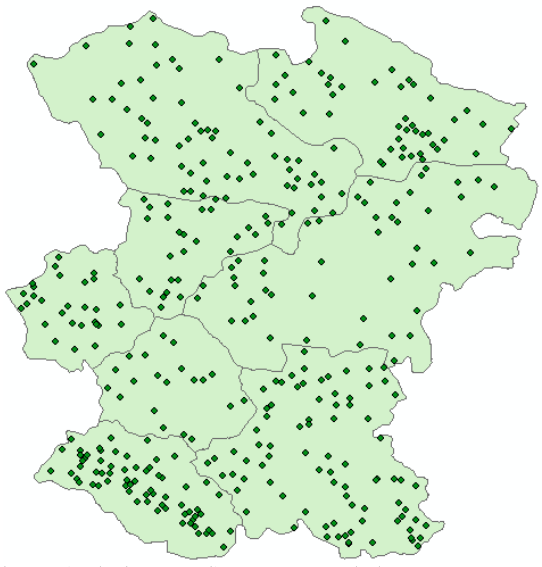
یک Itemset مجموعه‌ای است که بوسیله تعدادی عنصر ساخته می‌شود. در یک پایگاه داده رابطه‌ای که شامل m سطر و n ستون است، $(d_1, d_2, \dots, d_n)$ ابعاد مجموعه داده‌ها هستند که پیش‌بین نیز نامیده می‌شوند و d_j نشان دهنده j–امین بعد مجموعه داده‌ها است. $(v_{i1}, v_{i2}, \dots, v_{in})$ نشان دهنده مقادیر پیش‌بین‌های مختلف برای i–امین سطر از مجموعه داده‌ها است که v_{ij} مقدار j–امن بعد در i–امین سطر را نشان می‌دهد (۱۹).

در الگوریتم کاوش قوانین وابستگی چندبعدی برای تعیین وابستگی میان داده‌های چندبعدی، در مرحله اول باید در مورد ابعادی که مقادیر عددی دارند، گسسته‌سازی انجام گیرد. همچنین مقادیر عددی به دسته‌های گسسته تقسیم شده و تبدیل به مقادیر کیفی شوند. سپس تولید itemset های تکراری صورت می‌گیرد. این 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 عناصری با بعد یکسان باشد، آن کاندیدا حذف خواهد شد. مرحله هرس نیز مشابه الگوریتم پیشینی 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 (۲۰).

یافته‌ها

در این پژوهش به بررسی وابستگی میان بیماری تب‌مالت در انسان و دام و همچنین وابستگی بیماری با عوامل محیطی از قبیل دما، رطوبت، ارتفاع، نوع پوشش زمین و همچنین عوامل انسانی مانند شغل افراد، سابقه تماس آنها با دام و سابقه استفاده از لبنیات غیرپاستوریزه پرداخته شده است. ابتدا با استفاده از داده‌های توصیفی و روش‌های متداول، نقشه‌های مورد نیاز برای هر کدام از عوامل موثر تولید شده و پس از هم‌پوشانی نقشه‌ها، همبستگی بین نقشه‌ها با وقوع بیماری مشخص و جهت استخراج وابستگی‌های ناشناخته بین پارامترهای ذکرشده، از روش کاوش قوانین وابستگی استفاده گردید. نتایج بدست آمده جهت کشف دانش نشان‌دهنده ارتباط وقوع بیماری تب مالت با دما، ارتفاع، رطوبت، کاربری زمین، شغل بیمار، سابقه تماس با دام و سابقه استفاده از لبنیات غیرپاستوریزه می‌باشد.

برای تهیه داده‌های مکانی افراد مبتلا به بیماری تب مالت از آدرس آنها استفاده شده است، بطوریکه موقعیت مکانی بیماران در روستاها همان موقعیت مکانی روستاها در نظر گرفته‌شده و در شهرها با استفاده از نرم‌افزار گوگل ارث، موقعیت مکانی بیماران براساس آدرس آنها مشخص و در نهایت نقشه موقعیت مکانی بیماران مبتلا به بیماری تب مالت تهیه شده است (شکل ۱).



شکل ۱. نقشه موقعیت افراد مبتلا به تب مالت در سطح استان همدان

با توجه به اطلاعات دریافت شده از معاونت بهداشتی استان همدان، نقشه موقعیت افراد بیمار، نقشه سابقه تماس با دام (شکل ۲–الف)، نقشه سابقه استفاده از لبنیات غیرپاستوریزه (شکل ۲–ب) و نقشه شغل (شکل ۲–پ) در محیط نرم‌افزار ArcGIS تهیه شده است.

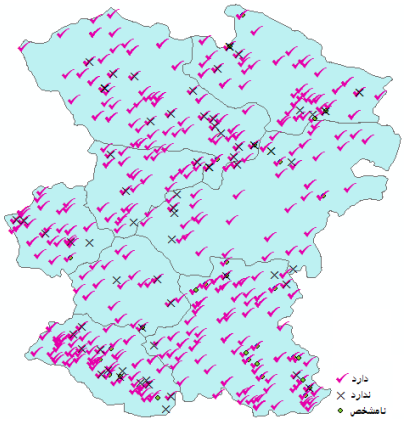
در این مطالعه جهت نمایش شدت پراکندگی افراد مبتلا به تب مالت در بازه زمانی یک سال (شکل ۲–ت) از روش تراکم

کرنل در محیط نرم‌افزار ArcGIS استفاده شده است که نتایج آن در پنج کلاس (تراکم خیلی کم، کم، متوسط، زیاد و خیلی زیاد) طبقه‌بندی شده است.

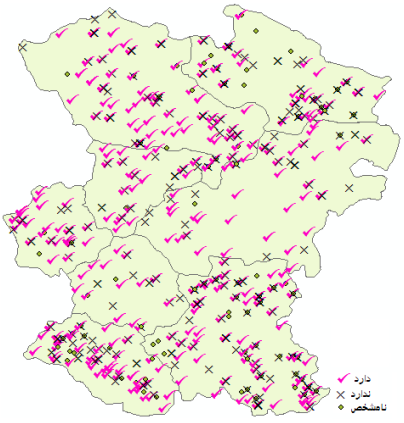
جهت بدست آوردن نقشه هم‌دما، با توجه به اینکه دما از جمله کمیت‌هایی است که همبستگی زیادی با ارتفاع دارد، بطوریکه با افزایش ارتفاع، مقدار دما کاهش می‌یابد. در منطقه مورد مطالعه، ارتفاع به عنوان عامل اصلی تاثیرگذار بر دما در نظر گرفته‌شد. برای شناخت تغییرات مکانی دما و تهیه نقشه‌های هم‌دما، نقشه دمای متوسط سالانه (شکل ۲–ث) با در نظر گرفتن همبستگی دما با ارتفاع برای منطقه مورد مطالعه بدست آمد.

برای تهیه نقشه شدت رطوبت نیز با واردکردن موقعیت مکانی ایستگاه‌های هواشناسی و داده‌های توصیفی رطوبت در محیط نرم‌افزار ArcGIS، با استفاده از روش درونیابی کریجینگ نقشه‌های شدت رطوبت (شکل ۲–ج) برای محدوده استان همدان تهیه شد.

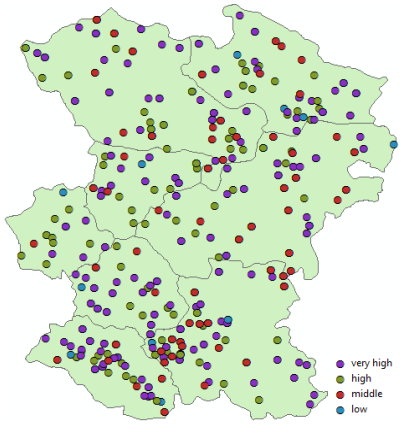
برای تهیه مدل رقومی زمین (شکل ۲–چ) از داده‌های SRTM و برای نقشه‌کاربری زمین (شکل ۲–ح) از نقشه‌های موجود در کشور استفاده شده است.



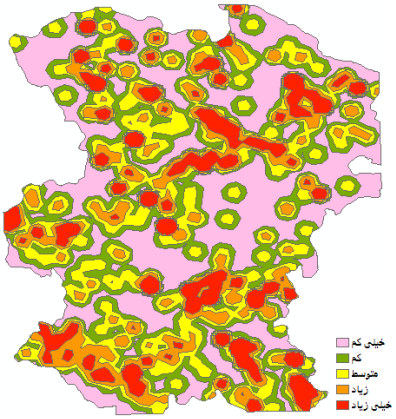
شکل ۲–الف. نقشه سابقه تماس با دام



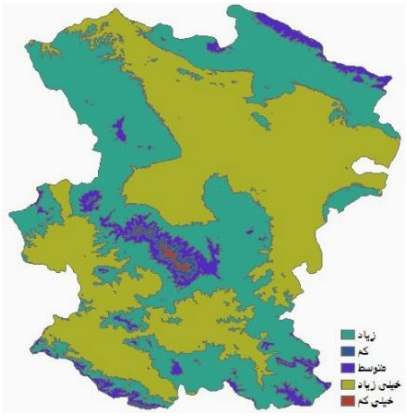
شکل ۲–ب. نقشه سابقه استفاده از لبنیات غیرپاستوریزه



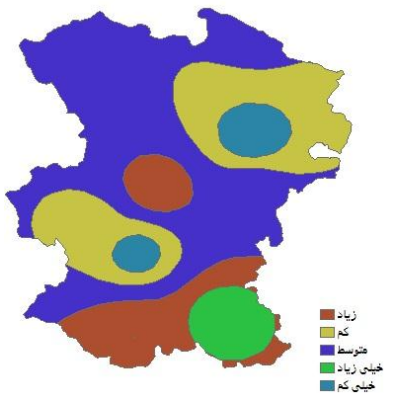
شکل ۲–پ. نقشه شغل



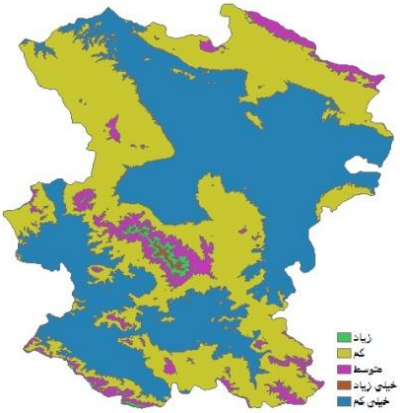
شکل ۲–ت. نقشه تراکم افراد مبتلا به تب‌مالت



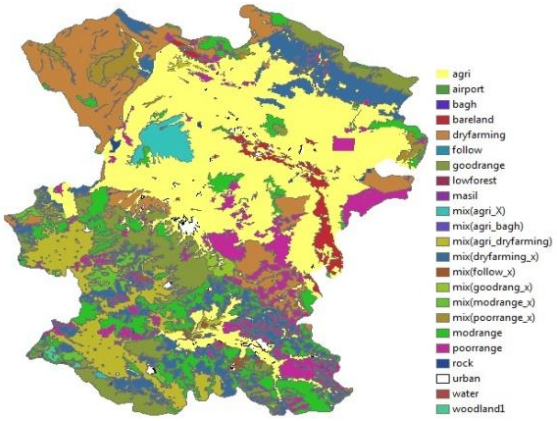
شکل ۲–ث. نقشه هم‌دما



شکل ۲–ج. نقشه رطوبت



شکل ۲–چ. نقشه مدل ارتفاعی رقومی زمین



شکل ۲–ح. نقشه کاربری زمین

پس از تهیه لایه‌های مکانی (نقشه‌ها)، همپوشانی لایه‌های مکانی انجام شده است. با همپوشانی لایه‌های مکانی شامل نقشه‌های تراکم بیماری، رطوبت، هم‌دما، مدل ارتفاعی رقومی و کاربری زمین همراه با درنظرگرفتن داده‌های انسانی شامل سابقه تماس با دام، سابقه استفاده از لبنیات غیرپاستوریزه و شغل بیمار در نرم‌افزار ArcGIS طبقه مربوط به هر داده برای هر بیمار مشخص گردیده است.

جدول (۱) قسمتی از جدول استخراج شده، بعد از همپوشانی لایه‌های مکانی را نشان می‌دهد. این جدول بیانگر یک جدول با چند بعد یا پیش‌بین است که به عنوان ورودی الگوریتم‌های پیشینی (Apriori) و کاوش قوانین وابستگی چندبعدی در نرم‌افزار SPSS استفاده شده است.

جدول۱. قسمتی از جدول چندبعدی استفاده شده به عنوان ورودی الگوریتم

شماره	تعداد افراد بیمار	دمای متوسط	ارتفاع	رطوبت متوسط	سابقه تماس با دام	سابقه استفاده از لبنیات غیرپاستوریزه	شغل
۱	خیلی زیاد	زیاد	کم	زیاد	دارد	دارد	دامدار–کشاورز
۲	خیلی زیاد	زیاد	کم	خیلی زیاد	دارد	دارد	دامدار–کشاورز
۳	خیلی زیاد	زیاد	کم	خیلی زیاد	دارد	ندارد	خانه دار
۴	خیلی زیاد	زیاد	کم	خیلی زیاد	دارد	دارد	کشاورز
۵	خیلی زیاد	زیاد	کم	خیلی زیاد	دارد	دارد	کودک
۶	خیلی زیاد	زیاد	کم	زیاد	دارد	دارد	دانش آموز
۷	خیلی زیاد	زیاد	کم	زیاد	دارد	دارد	خانه دار

با وارد کردن داده‌ها به الگوریتم کشف قوانین وابستگی چندبعدی و با تعریف حدآستانه حداقل پشتیبان ۱۰ درصد و حداقل اطمینان ۶۰ درصد، قوانین وابستگی موجود در داده‌ها کشف گردید. در اینجا هدف کشف قوانینی است که طرف دوم قانون، بیانگر میزان وقوع بیماری تب‌مالت در انسان باشند، مانند: (زیاد، افراد مبتلا) ⟶((متوسط، رطوبت) ∧ (خیلی زیاد، دما))

بنابراین این‌گونه قوانین، حفظ شده و بقیه قوانین حذف گردیدند. میزان پشتیبان قوانین کشف شده، بیانگر درصد وقوع تب‌مالت به همراه دسته‌هایی است که در قانون می‌آیند و میزان اطمینان قوانین کشف شده، بیانگر درصد وقوع تب‌مالت در دسته‌هایی است که در قانون می‌آیند. در نهایت، تعداد ۱۳۱ قانون کشف شد. جدول(۲) برخی از قوانین کشف شده را نشان می‌دهد.

جدول۲. قسمتی از قوانین کشف شده توسط الگوریتم

اطمینان(درصد)	پشتیبان(درصد)	قانون
۹۸/۴۹	۳۲/۳	(خیلی زیاد، افراد مبتلا)⟵((کم، ارتفاع)∧ (زیاد، دما)∧ (سابقه استفاده از لبنیات غیر پاستوریزه، دارد))
۹۸/۳۲	۱۵/۰۶	(خیلی زیاد، افراد مبتلا) ⟵((کم، ارتفاع) ∧ (زیاد، دما)∧ (دامدار_کشاورز، شغل)∧ (سابقه استفاده از لبنیات غیر پاستوریزه، دارد))
۹۸/۱۷	۱۷/۴۳	(خیلی زیاد، افراد مبتلا) ⟵((کم، ارتفاع) ∧(دارد، سابقه تماس با دام)∧ (زیاد، دما)∧ (متوسط، رطوبت)∧ (سابقه استفاده از لبنیات غیر پاستوریزه، دارد))
۹۸/۱۵	۱۳/۸۱	(خیلی زیاد، افراد مبتلا) ⟵((دامدار_کشاورز ، شغل) ∧(دارد، سابقه تماس با دام)∧ (متوسط، رطوبت))
۹۷/۸۶	۱۴/۳۳	(خیلی زیاد، افراد مبتلا) ⟵((خانه دار ، شغل) ∧(دارد، سابقه تماس با دام)∧ (کم، ارتفاع))
۹۷/۷۸	۸۶/۶۶	(خیلی زیاد، افراد مبتلا)⟵ (دارد، سابقه تماس با دام)
۹۶/۸۷	۱۱/۸۶	(زیاد، افراد مبتلا) ⟵ ⟵((دارد، سابقه تماس با دام)∧ (زیاد، دما)∧ (متوسط، رطوبت)∧ (سابقه استفاده از لبنیات غیر پاستوریزه، دارد))

به عنوان مثال قانون (خیلی زیاد، افراد مبتلا)⟵((کم، ارتفاع)∧ (زیاد، دما)∧ (سابقه استفاده از لبنیات غیر پاستوریزه، دارد)) با میزان پشتیبان۳۲/۳ درصد و اطمینان ۴۹/۹۸ درصد بیانگر این است که ۳۲/۳ درصد از افراد مبتلا به تب مالت در مکان‌هایی سکونت داشته‌اند که میزان دما در وضعیت زیاد و در ارتفاع کم بوده‌است و همچنین دارای سابقه استفاده از لبنیات غیر پاستوریزه نیز بوده‌اند. با توجه به قوانین کشف شده، قوانینی که دارای این شرایط هستند، دارای اطمینان بالاتری می‌باشند. پارامتر کاربری زمین به علت بالا بردن بسیار زیاد حجم اطلاعات، به ناچار به طور جداگانه بررسی شد و در نتیجه مناطق با کاربری زراعت دیم بیشترین آمار افراد مبتلا را با درصد پشتیبان ۲۵ درصد و اطمینان ۹۵ درصد داشته‌اند. در نتایج حاصل از عوامل غیرمکانی، در کلیه قوانین بدست آمده، مجموعا در ۵۳ قانون افراد دارای سابقه تماس با دام بوده‌اند و در هیچ قانون وابستگی خلاف این امر یافت نشد. فاکتور بعدی از عوامل غیرمکانی، فاکتور سابقه استفاده از لبنیات غیرپاستوریزه بوده است که در مجموع ۳۲ قانون مثبت و ۱۸ قانون منفی بدست آمد که از این امر می‌توان به این نتیجه دست یافت که این فاکتور نیز نقش نسبتا قابل توجهی را در شیوع بیماری داشته‌است. همچنین عامل شغل نیز در قوانین بدست آمده بیانگر این حقیقت بوده که افراد دارای شغل کشاورزی و دامداری دارای بیشترین درصد ابتلا بوده‌اند. بعد از این افراد، زنان خانه‌دار در قوانین کشف شده بیشترین آمار را به خود اختصاص داده‌اند.

بحث

تب مالت شایع‌ترین بیماری مشترک انسان و دام در ایران است، لذا کنترل این بیماری با شناسایی عوامل موثر در وقوع آن می‌تواند تاثیر مهمی بر سلامت انسان و صنعت دامداری خصوصا از نظر اقتصادی داشته باشد. تحقیقات نشان می‌دهد که محیط و مکان نقش مهمی در وقوع این بیماری دارند. در این تحقیق ابتدا داده‌های مکانی و توصیفی مورد نیاز برای عوامل محیطی مانند دما، رطوبت، ارتفاع، نوع پوشش زمین و عوامل انسانی مانند شغل و سابقه تماس بیمار با دام و سابقه استفاده از لبنیات غیرپاستوریزه تهیه شده و سپس نقشه‌های و لایه‌های مکانی وقوع بیماری و سایر لایه‌های عوامل محیطی و انسانی در محیط سیستم اطلاعات مکانی همپوشانی و تلفیق شده و نقشه همبستگی بیماری و اطلاعات وابسته مربوط به هر بیمار تهیه شده است. برای کشف قوانین وابستگی از روش الگوریتم کاوش قوانین وابستگی چندبعدی استفاده گردید. نتایج نشان می‌دهد که تمام عوامل می‌توانند در وقوع بیماری تب‌مالت موثر باشند.

نتایج قوانین کشف شده نشان می‌دهد که این بیماری از لحاظ عوامل محیطی در مناطق دارای دمای بالاتر، بیشتر اتفاق می‌افتد که با نتایج تحقیق محمدی و همکاران مطابقت دارد (۱۳). در این راستا پژوهش انجام شده توسط پاکزاد و همکاران جهت بررسی توزیع مکانی – زمانی وقوع بیماری تب‌مالت در ایران برای سال‌های ۲۰۱۱ تا ۲۰۱۴ در محیط سیستم اطلاعات جغرافیایی نشان می‌دهد که در اواسط فصل تابستان بیشترین و در اواسط فصل زمستان کمترین وقوع بیماری تب‌مالت اتفاق

افتاده است که با نتایج این پژوهش که وقوع تب‌مالت با افزایش دما رابطه مستقیم دارد، مطابقت دارد (۲۱). همچنین نتایج پژوهش عبادی فرد آذر و همکاران نشان می‌دهد که بیشترین شیوع بیماری تب مالت در فصل تابستان است (۲۲).

در نتایج حاصل از عوامل غیرمکانی، عامل شغل در قوانین بدست آمده بیانگر این حقیقت است که افراد دارای شغل کشاورزی و دامداری دارای بیشترین درصد ابتلا بوده‌اند. همچنین در افرادی که دارای سابقه تماس با دام یا استفاده از مواد لبنیاتی غیرپاستوریزه بوده‌اند، احتمال ابتلا به بیماری تب-مالت بیشتر است. به عبارت دیگر از لحاظ عوامل انسانی سابقه تماس با دام و مشاغل کشاورزی و دامداری دارای بالاترین تاثیر در وقوع بیماری بوده‌اند که با تحقیقات زینالی و همکاران و عبادی فرد آذر و همکاران مطابقت دارد (۲۲،۶). در این راستا نتایج پژوهش کنعانی نژاد و همکاران جهت بررسی اثر جمعیت و عوامل اکولوژیکی بر وقوع بیماری تب‌مالت در استان کهکیلویه و بویراحمد در محیط سیستم اطلاعات جغرافیایی نشان می‌دهد که توزیع بیماری تب‌مالت تحت تاثیر جمعیت انسانی، تراکم دام و عوامل اقلیمی و محیطی می‌باشد و در مناطق روستایی تراکم وقوع بیماری بیشتر است که با توجه به اینکه شغل اکثر روستائیان دامداری و کشاورزی می‌باشد، با نتایج این پژوهش، مطابقت دارد (۲۳). همچنین نتایج پژوهش Jun-Sik و همکاران در کره جنوبی نشان می‌دهد که در مناطق روستایی خطر ابتلا به بیماری تب‌مالت بیشتر است (۲۴).

همچنین در بیشتر قوانین بدست آمده عامل رطوبت دارای مقدار متوسطی بوده‌است، که وابسته به شرایط اقلیمی منطقه مورد مطالعه می‌باشد و بررسی دقیق تاثیر این پارامتر نیازمند انجام مطالعه در کل کشور با شرایط اقلیمی متفاوت می‌باشد. برای فاکتور نوع کاربری زمین نیز قوانین کشف شده نشان می-دهند که کاربری زراعت دیم دارای آمار افراد بیمار بیشتری بوده است. با توجه به اینکه معمولا پرورش دام در کنار مزارع کشاورزی انجام می‌شود و زمین‌های کشاورزی دیم مانند محصول گندم پس از برداشت برای چرای دام استفاده می‌شوند و در منطقه مورد مطالعه کشت دیم بیشترین مساحت کاشت محصولات کشاورزی را به خود اختصاص داده است، بیشترین وقوع بیماری در این کاربری اتفاق افتاده است که با تحقیقات کنعانی نژاد و همکاران مبنی بر اینکه در مناطق با پوشش گیاهی کشاورزی دیم و آبی و مرتع با پوشش ضعیف امکان وقوع بیماری تب‌مالت بیشتر است، هماهنگی دارد (۲۳).

نتیجه‌گیری

نتایج حاصل از این مطالعه بیانگر این است که هم عوامل محیطی از جمله دما، ارتفاع منطقه، رطوبت، کاربری زمین و هم عوامل غیرمکانی از جمله سابقه استفاده از لبنیات غیر پاستوریزه، سابقه تماس افراد بیمار با دام‌های آلوده و شغل افراد مبتلا شده، می‌توانند در شیوع بیماری تب‌مالت به نوعی دخیل باشند. لذا در بررسی قواعد همبستگی کشف شده، نشان داده‌شد که عوامل دما و ارتفاع دارای تاثیرات زیادی در قوانین کشف شده بوده‌اند، به گونه‌ای که مناطق دارای دمای زیادتر و ارتفاع پایین‌تر نسبت به دیگر مناطق، دارای افراد مبتلا شده بیشتری بوده‌اند. در خصوص ارتفاع با توجه به اینکه بیشتر دامداران و کشاورزان در مناطق دشت زندگی می‌کنند، در نتیجه مناطق با ارتفاع پایین‌تر با توجه به تراکم جمعیت و دام، پتانسیل بالاتری جهت شیوع بیماری تب‌مالت دارند. همچنین افراد دارای شغل کشاورزی و دامداری در قوانین کشف شده بیشترین آمار مبتلایان را به خود اختصاص داده‌اند.

در صورتیکه محل وقوع بیماری با استفاده از سیستم‌های تعیین موقعیت ماهواره‌ای ثبت شود، امکان تجزیه و تحلیل دقیق‌تر فراهم شده و نتایج بهبود یافته و دارای صحت بالاتری خواهد بود. با این وجود در این تحقیق سعی گردیده مکان وقوع بیماری تا حد امکان بر اساس آدرس بیمار دقیق ثبت شود و پارامترهای محیطی و انسانی با هم پردازش شوند. دانش کشف شده بیانگر این است که افرادی که دارای شغل وابسته به دام بوده و در محل‌هایی با ارتفاع کم، رطوبت متوسط و دیم‌زارها زندگی می‌کنند، در فصول گرم سال بیشتر در معرض خطر ابتلا به بیماری تب‌مالت قرار دارند. در این راستا می‌توان با شناسایی دام‌های مبتلا از تماس افراد با آن‌ها در راستای کاهش وقوع بیماری اقدام نمود.

نتایج کاربردی پژوهش

در این تحقیق پارامترهای انسانی و محیطی موثر در نحوه شیوع و وقوع بیماری تب مالت به همراه نقش و تاثیر آن‌ها با استفاده از سیستم اطلاعات مکانی و داده‌کاوی ویژگی‌های مکانی بررسی گردیده است که منجر به کشف قوانین مکانی بین پارامترهای مکانی محیطی و انسانی و وقوع بیماری تب مالت شده است، نتایج این تحقیق می‌تواند در برنامه‌ریزی، مدیریت و پیشگیری از وقوع بیماری تب مالت استفاده شده و در بهبود سلامت افراد و دام و در نتیجه کاهش هزینه‌های درمانی موثر باشد.

ملاحظات اخلاقی

در این پژوهش تمامی ملاحضات اخلاقی توسط نویسندگان رعایت شده است.

تضاد منافع

بدینوسیله نویسندگان اعلام می‌دارند که این اثر پژوهشی مستقل بوده و هیچ‌گونه تضاد منافعی برای نویسندگان این مقاله وجود ندارد.

تقدیر و تشکر

بدینوسیله از همکاری معاونت بهداشتی علوم پزشکی، مرکز تحقیقات بروسلوز، سازمان نقشه‌برداری و سازمان هواشناسی استان همدان به خاطر قراردادن داده‌های مورد نیاز تشکر و قدردانی می‌شود. این مقاله از پایان‌نامه کارشناسی ارشد خانم مهندس سیما قلی‌پور در موسسه آموزش عالی عمران و توسعه همدان با کد ۲۴۱۲۵۹۰ استخراج شده است.

References

1. Akbarmehr J. The prevalence of Brucella abortus and Brucella melitensis in local cheese produced in Sarab city, Iran and its public health implication. African Journal of Microbiology Research. 2011;5(12):1500-3. doi:10.5897/AJMR11.002
2. Beheshti S, Rezaian GR, Azad F, Faghiri Z, Taheri F. Seroprevalence of brucellosis and risk factors related to high risk occupational groups in Kazeroon, South of Iran. Int J Occup Environ Med. 2010;1(2):62-8.
3. Sathyanarayanan V, Razak A, Saravu K, Ananthakrishna SB, Mukhyprana Prabhu M, Vandana KE. Clinical profile of brucellosis from a tertiary care center in southern India. Asian Pac J Trop Med. 2011;4(5):397-400. doi:10.1016/s1995-7645(11)60111-9
4. Mugahi S, Nashibi R, Alavi SM, Gharkholu S, Najafi K. Epidemiological features, clinical manifestation and laboratory findings of patients with brucellosis. Arch Clin Infect Dis. 2014;9(1):e17270. doi:10.5812/archcid.17270
5. Hatami H. Brucellosis Epidemiology. The 2nd National Iranian Congress on Brucellosis; 2007 May 19-21; Shahid Beheshti University of Medical Sciences; p. 13-36.
6. Zeainali M, Shirzadi M, Hajrasouliha H. National manual for brucellosis control. Iran, Tehran: Raz Nahan: 2013. (Persian)
7. Berka P, Rauch J, Zighed D. Data mining and medical knowledge management. Hershey, Pa.: IGI Global; 2009.
8. Zhu L, Wu B, Cao C. [Introduction to medical data mining]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi.2003;20(3):559-62.
9. Cios KJ. Medical Data Mining and Knowledge Discovery. Physica-Verlag Heidelberg; 2000.
10. Han J, Kamber M, Pei J. Data Preprocessing. Data Mining: Concepts And Techniques. San Francisco: Morgan Kaufmann; 2006.
11. Hemalatha M, Saranya N N. A recent survey on knowledge discovery in spatial data mining. IJCI International Journal of Computer Science, 2011;8(3): 1-7.
12. Abdullayev R, Kracalik I, Ismayilova R, Ustun N, Talibzade A, Blackburn JK. Analyzing the spatial and temporal distribution of human brucellosis in Azerbaijan (1995 - 2009) using spatial and spatio-temporal statistics. BMC Infect Dis. 2012;12:185-. doi:10.1186/1471-2334-12-185
13. Mohammadi M, Taqian A, Eskandari M. Brucellosis disease zoning in Isfahan province. [Dissertation]. Isfahan University; 2013. (Persian)
14. Silverman B W. Density estimation for statistics and data analysis. CRC press; 1986.
15. Stein A. Universal kriging and cokriging as regression procedures. Corsten Biometrics, 1991;47:575 -587.
16. Jassar K K, Dhindsa KS. Comparative study of spatial data mining techniques. International Journal of Computer Applications. 2015.
17. Hayati Z, Sadeghi MM, Jafari N. Discovering the movement of electronic information users using the dependency law algorithm in data mining: a case study of the University of Utah University library, Australia. Journal of Library and Information Science, 2010;1(49):251-283. (Persian)
18. Cheung D W, Han J, Ng V, Wong C. Y. Maintenance Of Discovered Association Rules In Large Databases: An Incremental Updating Technique Conference Data Engineering (ICDE'96), New Orleans, LA, Feb. 1996:106–114.
19. Xu W X, Wang R J. A Fast Algorithm Of Mining Multidimensional Association Rules Frequently. 2006 International Conference on Machine Learning and Cybernetics. 2006:1199-1203.
20. Karimipour F, Kananisadat Y. Investigating the relation between prevalence of asthmatic allergy with the characteristics of the environment using fuzzy association rule mining. JGST. 2015;4 (3):117-130. (Persian)
21. Pakzad R, Pakzad I, Safiri S, Shirzadi MR, Mohammadpour M, Behroozi A, et al. Spatiotemporal analysis of brucellosis incidence in Iran from 2011 to 2014 using GIS. Int J Infect Dis. 2018;67:129-36. doi:10.1016/j.ijid.2017.10.017
22. Ebadifard Azar F, Jalilvand H, Abdi M, Nikuee D, Ghazaei SP, Pourrahimi M. Epidemiologic study of brucellosis disease in Divandarreh City of Kurdistan Province, Iran. Depiction of Health. 2019; 10(1):47-53. (Persian)
23. Kanannejad Z, Haghdoost A, Ghatee M.A, Azarifar F, Shahriari S, Moshfe A. Effect of human, livestock

population, climatic and environmental factors on the distribution of brucellosis in southwest Iran. Acta Medica Mediterranea, 2019, 35:22-59. doi:10.19193/0393-6384_2019_4_354

24. Lim J-S, Min K-D, Ryu S, Hwang S-S, Cho S-I. Spatial analysis to assess the relationship between human and bovine brucellosis in South Korea, 2005–2010. Sci Rep. 2019;9(1):6657. doi:10.1038/s41598-019-43043-7