

پیش‌بینی مکانی-زمانی مناطق پرخطر بیماری لپتوسپیروز با استفاده از روش‌های رگرسیون وزندار جغرافیایی و شبکه عصبی پرسپترون چندلایه

مهرداد آهنگرکانی^{۱*}، مهدی فرنقی^۲، محمدرضا شیرزادی^۳

^۱ دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه

نصیرالدین طوسی

mahangarcani@gmail.com

^۲ استادیار دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

farnaghi@kntu.ac.ir

^۳ رئیس مرکز کنترل و پیشگیری بیماری‌های واگیر وزارت بهداشت، درمان و آموزش پزشکی

shirzadim@gmail.com

(تاریخ دریافت خرداد ۱۳۹۵، تاریخ تصویب شهریور ۱۳۹۵)

چکیده

تشخیص عوامل بیماری‌زا، شناسایی تجمع مکانی بیماری و یافتن الگوی انتشار آن در محیط از ضروری‌ترین نیازها در زمینه بهداشت عمومی و مدیریت بیماری‌ها به شمار می‌آیند. بیماری لپتوسپیروز یکی از بیماری‌های مشترک انسان و دام می‌باشد که تقریباً در تمام نقاط جهان به‌ویژه در مناطق گرمسیری، نیمه‌گرمسیری و نواحی گرم و مرطوب شیوع بیشتری دارد. شرایط آب و هوایی معتدل و مرطوب در استانهای شمالی ایران این مناطق را در معرض خطر بیشتر برای ابتلا به این بیماری قرار داده است. اهداف اصلی این تحقیق بررسی سالانه توزیع مکانی بیماری لپتوسپیروز، کشف خوشه‌های مکانی و مکانی-زمانی بیماری و تهیه نقشه پیش‌بینی توزیع مکانی-زمانی آن در استانهای شمالی ایران می‌باشد. در این تحقیق داده‌های وقوع بیماری لپتوسپیروز به همراه داده‌های محیطی و توپوگرافی از قبیل میانگین دمای هوا، میانگین رطوبت، مجموع بارش سالیانه، ارتفاع، شیب، جهت شیب و تعداد روزهای یخبندان در سه استان گیلان، مازندران و گلستان از ابتدای سال ۱۳۸۹ تا پایان سال ۱۳۹۳ به صورت ماهانه جمع‌آوری شده‌اند. به منظور بررسی وجود و یا عدم وجود خودهمبستگی مکانی میان موارد وقوع بیماری لپتوسپیروز از شاخص‌های عمومی اندازه‌گیری خودهمبستگی مکانی همچون Moran's I و General G استفاده شده است. آزمون‌های خودهمبستگی عمومی برای هر ۶ سال به صورت جداگانه فرض تصادفی بودن توزیع بیماری در منطقه را رد کرده و بیانگر این است که توزیع بیماری در منطقه در هر ۶ سال خوشه‌ای می‌باشد. به منظور کشف خوشه‌های مکانی و شناسایی مناطق پرخطر بیماری لپتوسپیروز از شاخص‌های محلی کشف خوشه‌های مکانی از قبیل Local Moran's I و Local Gi استفاده شده است. در این تحقیق به منظور تهیه نقشه پیش‌بینی توزیع مکانی-زمانی بیماری لپتوسپیروز، از دو مدل رگرسیون وزندار جغرافیایی و شبکه عصبی پرسپترون چند لایه استفاده شده است. در نهایت معیارهای ضریب کاپا، MAPE، RMSE و R^2 جهت ارزیابی عملکرد مدلهای پیشنهادی به کار گرفته شده‌اند.

واژگان کلیدی: بیماری لپتوسپیروز، سامانه اطلاعات مکانی، توزیع مکانی-زمانی، رگرسیون وزندار جغرافیایی، شبکه عصبی، ایران

۱- مقدمه

لپتوسپیروز انسانی یک بیماری عفونی است که عامل ایجاد آن باکتری لپتوسپیرا می‌باشد و می‌تواند به صورت مستقیم و غیرمستقیم از حیوان به انسان منتقل شود [۱]. این بیماری یکی از گسترش‌یافته‌ترین بیماری مشترک انسان و حیوان به شمار می‌آید و به عنوان یک مشکل جهانی و مسئله‌ای برای سلامت و بهداشت عمومی در کشورهای در حال توسعه و توسعه یافته شناخته شده است [۲]. منبع اصلی این بیماری جوندگان و حیوانات وحشی می‌باشند که باکتری لپتوسپیرا را در فضولات خود دفع می‌کنند. برخی از حیوانات اهلی همچون سگ و گاو نیز در انتشار عفونت این بیماری مؤثرند و به عنوان منبع عفونت برای انسان و دام‌ها عمل می‌کنند [۳]. بیماری لپتوسپیروز در هر مکانی که انسان به طور مستقیم و یا غیرمستقیم در تماس با محیط آلوده باشد یافت می‌شود [۱]. گرچه این بیماری در اغلب مناطق جهان مشاهده شده است اما در مناطق گرمسیری پرباران، معتدل، مرطوب و دارای شرایط آب و خاک قلیایی شیوع بیشتری داشته و معمولاً در فصول گرم سال اتفاق می‌افتد [۴]. اگرچه تعداد افراد گرفتار به این بیماری دقیقاً مشخص نیست ولی گزارش‌های موجود نشان می‌دهد که میزان بروز بیماری در مناطق معتدل ۱-۰/۱ نفر در هر ۱۰۰,۰۰۰ نفر در سال و در مناطق مرطوب و گرمسیری میزان بروز ۱۰۰-۱۰ نفر در هر ۱۰۰,۰۰۰ نفر در سال می‌باشد [۵]. شیوع این بیماری در هر منطقه به وضعیت آب و هوایی، موقعیت جغرافیایی، جمعیت حیوانات اهلی و وحشی و در نهایت به وفور آب‌های سطحی بستگی دارد [۴]. در سال‌های اخیر سیر تحول بیماری لپتوسپیروز از یک بیماری سنتی شایع در میان کشاورزان، شالیکاران، ماهیگیران و به صورت کلی جوامع روستایی به یک بیماری عمومی و اپیدمی در جوامع شهری با سطح بهداشت نامناسب و ضعیف تبدیل شده است [۴]. با توجه به شرایط آب و هوایی معتدل، مرطوب، وضعیت جلگه‌ای و همچنین به دلیل داشتن شرایط آب و خاک قلیایی در استان‌های شمالی ایران، احتمال وقوع بیماری لپتوسپیروز در این مناطق بیشتر بوده و موارد قابل توجهی از بیماری انسانی در این مناطق گزارش شده است [۲]. در استان‌های شمالی و حاشیه دریای خزر فعالیت اصلی روستاییان کشت برنج می‌باشد و اغلب روستاییان نیز در منازل خود یک یا چند

رأس دام نگهداری می‌کنند و همچنین در اغلب روستاها از رودخانه و آب‌های سطحی برای آبیاری مزارع و زمین‌های کشاورزی استفاده می‌شود. مجموعه این شرایط و عوامل وضعیت مناسبی را برای انتشار و پخش بیماری لپتوسپیروز در استان‌های شمالی ایران ایجاد می‌کند [۴]. با توجه به بومی بودن این بیماری در مناطق شمالی ایران و افزایش شیوع بیماری در فصول گرم سال، همزمان با شروع فعالیت‌های کشاورزی می‌توان با تشخیص درست و به موقع این بیماری، میزان مرگ و میر ناشی از آن را به طور قابل توجهی کاهش داد [۲]. تشخیص عوامل بیماری‌زا، شناسایی تجمع مکانی بیماری، یافتن الگوی انتشار بیماری در محیط و کشف عوامل تشدید کننده آن از ضروری‌ترین نیازها در زمینه بهداشت عمومی و مدیریت بیماری‌ها به شمار می‌آیند [۶]. انتشار بیماری‌ها و مسائل بهداشت عمومی با جغرافیای منطقه رابطه مستقیم دارند و همواره دارای بعد مکانی می‌باشند [۷]. از طرفی با توجه به حجم بودن داده‌های مکانی و پیچیدگی‌های موجود در این قسم داده‌ها، سامانه اطلاعات مکانی (GIS) با قابلیت تجزیه و تحلیل حجم عظیمی از داده‌ها و اطلاعات مکان مرجع می‌تواند به عنوان ابزاری مفید و کارآمد در جهت مدیریت و تصمیم‌گیری بهینه در زمینه بهداشت عمومی در اختیار کارشناسان و تصمیم‌گیران قرار گیرد [۸]، [۹].

اهداف اصلی این تحقیق بررسی سالانه توزیع مکانی بیماری لپتوسپیروز، کشف خوشه‌های مکانی و مکانی-زمانی بیماری به منظور درک بهتر جنبه‌های همه-گیرشناسی^۱ مکانی-زمانی این بیماری و تهیه نقشه پیش-بینی توزیع مکانی-زمانی آن در سطح بخش‌های استان-های شمالی ایران می‌باشد. بر این اساس این مقاله در ۵ فصل به قرار زیر تدوین شده است. پس از مقدمه در فصل دوم، به بررسی مطالعات انجام شده در زمینه بیماری‌های واگیر و به طور خاص بیماری لپتوسپیروز و نقش GIS در مدلسازی و تهیه نقشه‌های پیش‌بینی و شناسایی مناطق پرخطر پرداخته شده است. در فصل سوم به مفاهیم پایه تجزیه و تحلیل‌های مکانی و آماری به منظور بررسی وجود خودهمبستگی مکانی بیماری و کشف خوشه‌های مکانی بیماری لپتوسپیروز و مدل‌های مورد استفاده در تحقیق پرداخته شده است. در فصل چهارم ضمن اشاره به منطقه

^۱ Epidemiology

مورد مطالعه و داده‌های مورد استفاده در تحقیق، نحوه مدل‌سازی بیماری با استفاده از مدل‌های مورد استفاده در تحقیق و به طور خاص نتایج و خروجی فصل سوم توصیف و تحلیل شده‌اند. در نهایت در فصل پنجم به جمع‌بندی مسائل مطرح در تحقیق اختصاص یافته و پیشنهادهای جهت ادامه تحقیق فوق در آینده ارائه شده است.

۲- پیشینه تحقیق

استفاده از بررسی‌های همه‌گیرشناسی برای حل مشکلات سلامت جامعه سابقه‌ای طولانی دارد. تاریخ دانش پزشکی انباشته از کشف راهکارهای فراوان مبارزه با بیماری‌هاست. بررسی‌های جان اسنو در اوایل نیمه دوم قرن نوزدهم در بریتانیا در زمینه چگونگی انتشار بیماری وبا در شهر لندن و کنترل آن و یا بررسی‌های ادوارد جنز در زمینه بیماری آبله انسانی و تهیه واکسن آن از نمونه‌های برجسته تاریخی این نوع پژوهش‌های همه‌گیرشناسی هستند ([۱۰]، [۱۱]). در سال‌های اخیر تلاش‌های فراوانی به منظور مطالعه بر روی انتشار بیماری‌های واگیر و چگونگی شیوع بیماری‌ها صورت گرفته است. تعدادی از مطالعات سراسر دنیا نشان داده‌اند که پارامترهای محیطی، آماری و جمعیت شناختی در رابطه با بوم‌شناسی^۱ بیماری‌های واگیر می‌توانند به عنوان عامل کلیدی و پایه به منظور توسعه مدل‌های مکانی و پیش‌بینی مناطق در معرض خطر بیماری ایفای نقش کنند ([۱۲]، [۶]). امروزه استفاده از سامانه‌های اطلاعات مکانی (GIS) به منظور مدیریت و برنامه‌ریزی مسائل بهداشتی و بیماری‌ها مورد توجه محققین قرار گرفته است و دریافته‌اند که استفاده از چنین سامانه‌هایی می‌تواند برای تولید نقشه‌های پیش‌بینی و شناسایی مناطق بومی بیماری‌های واگیر بکار گرفته شوند ([۱۳]، [۱۴]). به عنوان مثال، رجیبی و همکاران در سال ۲۰۱۲ میلادی روش‌های تصمیم‌گیری چندمعیاره همراه با GIS را با هدف شناسایی مناطق پرخطر بیماری کالآزار در شهرستان‌های شرق استان آذربایجان شرقی مورد استفاده قرار دادند [۱۵]. مطالعات بسیاری در رابطه با بیماری‌های واگیر و عفونی در قسمت‌های مختلف جهان انجام شده‌اند که در این مطالعات از روش‌های رگرسیون و شبکه عصبی

برای مدل‌سازی بیماری استفاده شده است. به عنوان نمونه می‌توان به مطالعات زیر اشاره نمود:

در مطالعه‌ای ریمود و همکاران در سال ۲۰۰۴ میلادی از مدل رگرسیون لجستیک با در نظر گرفتن متغیرهای محیطی و توپوگرافی برای پیش‌بینی بیماری سالک در کلمبیا استفاده نمودند و به این نتیجه رسیدند که دو متغیر ارتفاع و کاربری زمین به عنوان موثرترین متغیرها شناخته شده‌اند. در نهایت با استفاده از معیارهای عملکرد مدل نظیر PPV^۲، NPV^۳ و Kappa میزان دقیق بودن پیش‌بینی مدل بدست آمده را مورد بررسی قرار دادند [۱۶]. سدهی و همکاران در سال ۲۰۱۰ میلادی، در مطالعه‌ای توانایی مدل‌های شبکه عصبی مصنوعی با رگرسیون لجستیک را به منظور پیش‌بینی سندرم متابولیک در نمونه‌ای از افراد شرکت کننده در مطالعه قند و لیپید تهران مقایسه نمودند. ایشان از تحلیل منحنی راک (ROC^۴) به منظور مقایسه قدرت پیش‌بینی مدل‌ها استفاده کردند. مساحت زیر منحنی راک برای مدل‌های رگرسیون لجستیک و شبکه عصبی به ترتیب ۰/۷۳۹ و ۰/۸۹۰ بدست آمد. این نتیجه نشان داد که مدل شبکه عصبی نسبت به مدل رگرسیون لجستیک از دقت بیشتری برای پیش‌بینی سندرم متابولیک در افراد مورد بررسی برخوردار است [۱۷]. در مطالعه‌ای، لین و ون در سال ۲۰۱۱ میلادی، به بررسی تغییرات محلی بیماری تب دانگ و پشه‌های ناقل این بیماری با استفاده از روش رگرسیون وزندار جغرافیایی^۵ و مقایسه آن با روش کمترین مربعات معمولی^۶ پرداختند. معیار ارزیابی R² بدست آمده برای این دو روش به ترتیب برابر با ۰/۵۹ و ۰/۰۴ بدست آمد که به خوبی وجود تغییرات محلی و برتری روش GWR را نسبت به روش OLS نشان می‌دهد [۱۸]. همچنین، علی‌اکبرپور و همکاران در سال ۲۰۱۲ میلادی، به شناسایی عوامل آب و هوایی و محیطی موثر بر بیماری سالک در استان فارس با استفاده از مدل رگرسیون خطی و مدل رگرسیون وزندار جغرافیایی پرداختند. نتایج حاصل از مدل‌سازی نشان دادند که مدل رگرسیون وزندار جغرافیایی به دلیل ویژگی خودهمبستگی و عدم ایستایی

^۲ Positive Predictive Values

^۳ Negative Predictive Value

^۴ Receiver Operating Characteristic

^۵ Geographically Weighted Regression (GWR)

^۶ Ordinary Least Squares

^۱ Ecology

در داده‌های مکانی، نسبت به مدل رگرسیون خطی عملکرد بهتری داشته است. همچنین براساس رگرسیون وزندار جغرافیایی، تعداد روزهای بارانی، حداقل دما، شدت باد و حداکثر رطوبت نسبی به عنوان متغیرهای مهم در مدل‌سازی بیماری شناخته شده‌اند [۱۹]. زاده و همکاران در سال ۲۰۱۲ میلادی، به تشخیص بیمای سرطان سینه با استفاده از ترکیب شبکه عصبی و الگوریتم ژنتیک پرداختند [۲۰]. ریس‌مالا و همکاران در سال ۲۰۱۳ میلادی، بیماری مالاریا را در کشور اندونزی مورد بررسی قرار دارند. در این مطالعه از روش شبکه عصبی مصنوعی همراه با الگوریتم ژنتیک و داده‌های زیست‌محیطی و هواشناسی استفاده کردند. نتایج بدست آمده کارایی بالای این روش در پیش‌بینی بیماری مالاریا در سال ۲۰۰۸ و ۲۰۰۹ با ۷۵ درصد دقت و وجود همبستگی کافی بین بیماری و متغیرهای محیطی را نشان داد [۲۱]. در مطالعه‌ای، رجبی و همکاران در سال ۲۰۱۴ میلادی، از مدل شبکه عصبی RBFLN به منظور تهیه نقشه آسیب‌پذیری بیماری کال‌آزار در استان آذربایجان شرقی استفاده نمودند. در این تحقیق متغیرهای محیطی و توپوگرافی به عنوان عوامل موثر بر شیوع بیماری در نظر گرفته شدند. نقشه آسیب‌پذیری بدست آمده از مدل شبکه عصبی نشان داد که حاشیه رودخانه، مناطق کم ارتفاع و با پوشش گیاهی کم و پراکنده و روستاهای عشایرنشین بدون مراکز درمانی و بهداشتی واقع در منطقه مورد مطالعه، از مکان‌های بسیار مناسب برای رشد ناقلین و مخازن بیماری بوده و این مناطق در معرض خطر بیشتری برای شیوع بیماری کال‌آزار محسوب می‌شوند [۲۲]. همچنین اخوان و همکاران در سال ۲۰۱۴، از شبکه عصبی پرسپترون چندلایه با هدف تهیه نقشه آسیب‌پذیری بیماری سالک در استان ایلام استفاده نمودند. در این تحقیق نتایج حاصل از اولویت‌دهی متغیرهای موثر بر شیوع بیماری سالک توسط ضریب همبستگی نشان دادند که عوامل رطوبت و ارتفاع دارای بیشترین اهمیت در تهیه نقشه آسیب‌پذیری بیماری می‌باشند. نقشه آسیب‌پذیری بدست آمده بیانگر این نکته بود که مناطق کم ارتفاع و دارای میانگین دمای بالا، مکان‌های مناسب برای شیوع بیماری سالک می‌باشند [۲۳]. وانگ و همکاران در سال ۲۰۱۵ میلادی، به پیش‌بینی تعداد بیماران یک نوع اسهال واگیر تحت تاثیر متغیرهای اقلیمی و آب و هوایی در چین

پرداختند. به این منظور، ایشان از یک شبکه عصبی سه لایه استفاده کردند که با الگوریتم لوبنرگ-مارکوات آموزش داده شده بود و نتایج آن را با سه روش رگرسیونی SVR، RFR و MLR مقایسه نمودند. ارزیابی‌ها بر اساس داده‌های واقعی، حاکی از برتری شبکه عصبی نسبت به سه روش رگرسیونی بود [۲۴].

مطالعات بسیاری در سال‌های گذشته در رابطه با بیماری لپتوسپیروز در سرتاسر جهان گزارش شده است که تعداد بسیار زیادی از این مطالعات به جنبه‌های پزشکی و نوع دامی بیماری لپتوسپیروز پرداخته‌اند ([۳]، [۲۵]، [۲۶]، [۲۷]). اما در خصوص بیماری لپتوسپیروز انسانی و تحلیل مکانی و مدل‌سازی در مورد آن تاکنون مطالعات محدودی سراسر دنیا صورت گرفته است. در سال ۲۰۱۴ میلادی، وگا کوردور و همکاران به بررسی بیماری لپتوسپیروز در جزیره ترینیداد از کشورهای حوزه دریای کارائیب پرداختند. ایشان از روش رگرسیون پواسن وزندار جغرافیایی (GWPR) برای مدل‌سازی بیماری لپتوسپیروز استفاده نمودند. در این مطالعه با استفاده از سامانه اطلاعات مکانی و تصاویر سنجنده MODIS متغیرهایی همچون پوشش زمین، شاخص رطوبت^۱، درصد خاک‌های زهکشی، توپوگرافی، متوسط بارش سالیانه و غیره استخراج و در مدل‌سازی استفاده شدند. نتایج این تحقیق مشخص نمود که متغیرهای بارش، توپوگرافی و شاخص رطوبت بیشترین تاثیر را بر شیوع بیماری لپتوسپیروز دارا می‌باشند [۱]. همچنین، مطالعه‌ای در سال ۲۰۱۵ میلادی توسط نیا و همکاران در استان گیلان صورت گرفته است. در این مطالعه به منظور بررسی اثر متغیرهای محیطی موثر بر شیوع بیماری لپتوسپیروز و کشف خوشه‌های مکانی ممکن، توزیع سالانه مکانی و آماری بیماری لپتوسپیروز را تجزیه و تحلیل کرده‌اند. نتایج حاصل از این مطالعه نشان دادند که وقوع بیماری عمدتاً در فصول گرم سال (از خرداد تا شهریور) همزمان با شروع فصل کشاورزی و کشت برنج اتفاق افتاده است. همچنین در نواحی مرکزی استان به دلیل داشتن زمین‌های کشاورزی بیشتری نسبت به نواحی دیگر، بیشترین شیوع بیماری دیده شده است [۲۸]. تمامی این مطالعات، بر تجزیه و تحلیل و بررسی ارتباط میان وقوع بیماری لپتوسپیروز و

^۱ WETINDEX

پارامترهای موثرش تمرکز کرده‌اند. از این‌رو، با توجه بررسی‌های صورت گرفته و طبق بهترین اطلاعات، توزیع مکانی- زمانی بیماری لپتوسپیروز در هیچ کدام یک از مطالعات پیشین بررسی و ارزیابی نشده است.

۳- مواد و روش‌ها

در این بخش روش‌های تحلیلی مورد استفاده در این پژوهش جهت تهیه نقشه پیش‌بینی مکانی- زمانی بیماری لپتوسپیروز تشریح می‌شوند. در این راستا ابتدا تحلیل‌های زمین آمار استفاده شده جهت بررسی وجود و یا عدم وجود خودهمبستگی مکانی میان موارد وقوع بیماری ارائه شده‌اند. سپس در ادامه دو روش رگرسیون وزندار جغرافیایی و شبکه عصبی پرسپترون چندلایه^۱ به عنوان مدل‌های پیشنهادی جهت پیش‌بینی توزیع مکانی- زمانی بیماری لپتوسپیروز تشریح شده‌اند.

۳-۱- تحلیل‌های زمین آمار

توزیع مکانی بسیاری از رویدادها از جمله بیماری‌ها، تصادفات، جرائم و غیره بیانگر حضور عوامل محیطی ایجاد کننده آنها می‌باشد [۹]. در مطالعات محیطی غالباً با داده‌هایی سر و کار داریم که مستقل از یکدیگر نمی‌باشند و به نوعی وابستگی آن‌ها ناشی از موقعیت و مکان قرار گرفتن داده‌ها در فضای مورد مطالعه می‌باشد [۲۹]. از طرفی فرض اصلی بیشتر روش‌های آماری معمول بر اساس استقلال داده‌ها می‌باشد. از این‌رو به دلیل وجود همبستگی و اثر مکانی میان این قسم از داده‌ها، این فرض در حقیقت محقق نشده و داده‌ها به هم وابسته هستند، لذا روش‌های آماری معمول نمی‌توانند روشی مناسب برای بررسی چنین داده‌هایی باشند ([۳۰]، [۳۱]، [۳۲]). بنابراین روش‌های زمین آمار به عنوان گزینه‌ای مناسب در تحلیل چنین داده‌هایی می‌توانند مورد استفاده قرار گیرند. بنابراین به منظور مدلسازی برخی از رویدادها همچون بیماری‌ها، لازم است که ابتدا خودهمبستگی مکانی بین موارد وقوع رویداد بررسی شود و تعیین گردد که الگوی مکانی رویداد موردنظر در منطقه از کدام توزیع (تصادفی، یکنواخت یا خوشه‌ای) پیروی می‌کند. در مورد

خودهمبستگی مکانی فرض صفر بیان می‌کند که هیچ الگوی مکانی میان موارد وقوع رویداد مورد نظر وجود ندارد. فرض صفر یک نوع فرض مورد استفاده در آمار است و بیانگر آن است که هیچ اهمیت آماری در یک مجموعه از مشاهدات داده وجود ندارد. فرض صفر تا زمانی پذیرفته است که شواهد آماری، آن را برای یک فرض جایگزین باطل نکنند [۳۳]. از این‌رو به منظور بررسی و تشخیص وجود خودهمبستگی مکانی، آزمون‌های مختلفی از قبیل شاخص Moran's I، شاخص متوسط فاصله از نزدیک‌ترین همسایه^۲، شاخص آمار اسکن مکانی، شاخص General G و غیره پیشنهاد شده‌اند [۳۴]. با توجه به اینکه واحد مکانی مورد نظر در این مطالعه به صورت چندضلعی می‌باشد، از شاخص‌های Moran's I و General G به دلیل کارایی بیشتر آنها نسبت به روش‌های دیگر در زمینه بررسی خودهمبستگی مکانی بین موارد وقوع بیماری‌ها در داده‌های چندضلعی استفاده شده است ([۳۵]، [۳۶]). در ادامه هر یک از این شاخص‌ها تشریح می‌شوند.

۳-۱-۱- شاخص Moran's I

این شاخص یکی از ابزارهای بررسی خودهمبستگی مکانی بین داده‌های توصیفی- عددی منتسب به چندضلعی‌های پوشش دهنده‌ی منطقه مطالعه، مانند تعداد موارد وقوع بیماری در محلات یک شهر، می‌باشد. در یک مجموعه داده، شاخص موران بین مقادیر -1 تا $+1$ محاسبه می‌شود. چنانچه مقدار شاخص I بزرگتر از صفر باشد خودهمبستگی مکانی مثبت، اگر مقدار I کمتر از صفر باشد خودهمبستگی مکانی منفی و اگر مقدار I نزدیک به صفر باشد به معنی عدم وجود خودهمبستگی مکانی خواهد بود.

هر چه مقدار عددی (قدر مطلق) این ضریب مقدار بالاتری داشته باشد، بیانگر تجمع زیاد و هر چه مقدار پایین تری داشته باشد بیانگر پراکندگی می‌باشد [۳۷]. مقدار این شاخص از رابطه (۱) تعیین می‌شود:

$$I = \left(\frac{N}{\sum_{i=1}^N \sum_{j=1}^N W_{ij}} \right) * \left(\frac{\sum_{i=1}^N \sum_{j=1}^N W_{ij} Z_i Z_j}{\sum_{i=1}^N (Z_i)^2} \right) \quad (1)$$

$$Z_i = (x_i - \bar{x})$$

$$Z_j = (x_j - \bar{x})$$

^۲ Average Nearest neighbor distance

^۱ Multilayer Perceptron Neural Network (MLP)

که در رابطه فوق داریم:

x_i و x_j : تعداد موارد وقوع رویداد در i امین و j امین چندضلعی.

x و N : تعداد متوسط موارد وقوع رویداد در چندضلعی‌ها و تعداد کل موارد وقوع رویداد در منطقه.

W_{ij} : وزن مکانی میان موارد i و j (نشانگر نزدیکی). در حالت کلی زمانی که دو چندضلعی i و j همسایه باشند $W_{ij} = 1$ و در غیر این صورت برابر صفر می‌باشد (در این مطالعه معیار وزندهی بر اساس همسایگی دو چندضلعی تعریف شده است). همچنین W_{ij} می‌تواند به صورت ماتریس وزن، فاصله مراکز چندضلعی‌ها، طول مرز مشترک و یا ترکیب طول مرز مشترک و فاصله چندضلعی‌ها نیز تعریف شود.

۳-۱-۲- شاخص General G

این شاخص در سال ۱۹۹۲ توسط Getis و Ord، به منظور بررسی تجمع مقادیر خیلی بزرگ یا خیلی کوچک یک رویداد (مانند بروز بیماری) که به ترتیب نشانگر نقاط داغ^۱ (مناطق بحرانی و پرخطر) و نقاط سرد^۲ (مناطق بدون (کم) خطر) هستند، توسعه یافت [۳۸]. در این شاخص مقادیر مثبت Z-score بیانگر نقاط داغ و مقادیر منفی Z-score نقاط سرد را نشان می‌دهد. در واقع Z-score امتیاز استاندارد شده‌ای است که میزان متمرکز بودن یا پخش‌شدگی بیماری را نشان می‌دهد. هر چه این مقدار از صفر فاصله بیشتری داشته باشد، بیانگر آن است که پدیده بیماری روند خوشه‌ای‌تر یا پراکنده‌تری دارد. اگر علامت Z مثبت باشد یعنی در بین مناطقی که بیماری مقدار بالایی دارد، روند خوشه‌ای وجود دارد و اگر منفی باشد یعنی در مناطقی که بیماری مقدار پایینی دارد، روند خوشه‌ای است [۳۹]. پارامترهای شاخص General G مشابه پارامترهای شاخص Moran's I بوده و مقدار این شاخص از رابطه (۲) محاسبه می‌شود:

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N W_{ij} x_i x_j}{\sum_{i=1}^N \sum_{j=1}^N x_i x_j}, \quad \square \quad j \neq i \quad (2)$$

۳-۲- رگرسیون وزندار جغرافیایی

روشهای رگرسیون از روشهایی هستند که جهت مدل‌سازی و کمی نمودن ارتباط میان یک متغیر وابسته و

یک یا چند متغیر مستقل به کار گرفته می‌شوند. روشهای رگرسیون معمولی یک رابطه ثابت و یکسان میان متغیرها در نظر می‌گیرند و اثر مکانی متغیرها را به حساب نمی‌آورند. برتری عمده روش رگرسیون وزندار جغرافیایی در مقابل روش رگرسیون معمولی توانایی آن در بررسی اثر مکانی متغیرهاست. به گونه‌ای که رابطه بین متغیرها بر حسب مکان تغییر می‌کند [۴۰]. در روش رگرسیون وزندار جغرافیایی، رابطه رگرسیونی برآورد شده در هر منطقه متفاوت از منطقه دیگر می‌باشد [۴۱]. GWR یک روش رگرسیون محلی است که به طور قابل توجهی رگرسیون معمولی را برای استفاده در داده‌های مکانی بهبود داده است. به طوریکه اگر براساس روش رگرسیون معمولی برای تمام مناطق از رابطه یکسانی استفاده شود، برآوردی غیر دقیق حاصل خواهد شد [۴۰]. وجود همبستگی مکانی میان داده‌های مکانی و اثر مکانی‌ای که هر داده بر داده‌های مجاور خود دارد، باعث می‌شوند که رابطه بین متغیرهای مستقل و وابسته در قسمتی از منطقه مورد مطالعه مثبت و در بخشی دیگر منفی شود. از این‌رو مدل‌های رگرسیون معمولی قادر به تشخیص رابطه دقیق میان متغیرهای مستقل و وابسته نمی‌باشند [۴۲]. بنابراین هنگام بررسی یک پدیده که در آن به دلیل وجود همبستگی و اثر مکانی میان مشاهدات، تغییرات محلی وجود دارد، استفاده از روش‌های معمولی نامناسب بوده و پاسخ خوبی برای مسئله نخواهند داشت. از این‌رو، لازم است روشی از قبیل رگرسیون وزندار جغرافیایی استفاده شود که به نحوی ساختار همبستگی داده‌ها در تحلیل آن‌ها لحاظ گردد. به منظور حل مسئله همبستگی میان داده‌های مکانی، در رگرسیون وزندار جغرافیایی هر داده مشاهداتی به نسبت فاصله‌اش از موقعیت مورد برآورد وزندهی می‌شوند و از این نظر دقت جغرافیایی کار بالا می‌رود [۴۳]. رابطه (۳) رگرسیون وزندار جغرافیایی را نشان می‌دهند [۴۲].

$$\beta(u) = (X^T W(u) X)^{-1} X^T W(u) Y \quad (3)$$

در رابطه فوق، Y مقدار متغیر وابسته در نقطه (مرکز چندضلعی) (u) ، X مقادیر مربوط به متغیر مستقل در همان نقطه (u) ، β برآوردگرهای مدل و W ماتریس مربعی وزنها و وابسته به موقعیت نقطه (u) در منطقه هستند. وزن مربوط به ضرایب هر نقطه کاملاً وابسته به موقعیت مکانی آن نقطه و همچنین موقعیت آن نسبت به سایر

^۱ Hot Spot

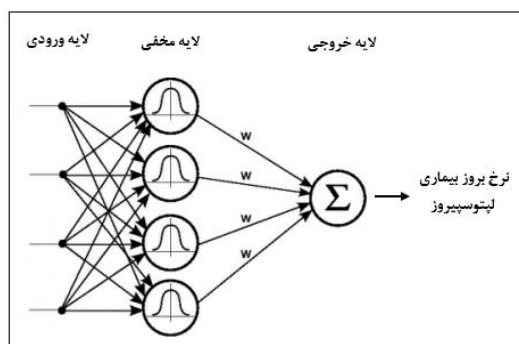
^۲ Cold Spot

استفاده قرار می‌گیرند. نحوه عمل پرسپترون چند لایه به این صورت است که الگویی به شبکه عرضه می‌شود و خروجی آن محاسبه می‌گردد. مقادیر خروجی واقعی و خروجی مطلوب باعث می‌گردد که ضرایب شبکه تغییر یابد؛ به گونه‌ای که در مراحل بعد خروجی صحیح‌تری حاصل شود. برای موفقیت در آموزش شبکه باید خروجی آن را به تدریج به خروجی مطلوب نزدیک کرد و میزان خطا را کاهش داد [۵۱]. معیار دقت در این نوع از شبکه‌ها، معمولاً RMSE^۱ (خطای جذر میانگین مربعات) و MAPE^۲ (درصد میانگین مطلق خطا) می‌باشند که از تفاوت میان مقدار پیش‌بینی شده توسط مدل و مقدار واقعی محاسبه می‌شوند [۵۲]، [۵۳]. این معیارها را می‌توان به صورت روابط (۴) و (۵) نشان داد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - y_i)^2}{N}} \quad (4)$$

$$MAPE = \frac{1}{N} * \sum_{i=1}^N \left| \frac{(Y_i - y_i)}{Y_i} \right| * 100 \quad (5)$$

که در این روابط: N ، تعداد پیش‌بینی‌ها، Y_i و y_i به ترتیب بیانگر مقادیر واقعی و پیش‌بینی شده می‌باشند.



شکل ۱- معماری شبکه عصبی چند لایه (اقتباس شده از [۵۱]).

۴- پیاده‌سازی و نتایج

در این قسمت از تحقیق ضمن اشاره به منطقه مورد مطالعه و داده‌های مورد استفاده، نحوه مدل‌سازی بیماری با استفاده از روش‌های رگرسیون وزندار جغرافیایی و شبکه عصبی پرسپترون چند لایه، نتایج حاصل از تجزیه و تحلیل‌های مکانی و آماری، نتایج بدست آمده از مدل‌ها و

نقاط در منطقه مورد مطالعه می‌باشد. این بدین معنا می‌باشد که مشاهدات نزدیک به نقطه برآوردی وزن و تاثیر بیشتری در برآورد دارا می‌باشند و مشاهدات دورتر دارای وزن و تاثیر کمتری هستند. در واقع موقعیت هر نقطه در منطقه مورد مطالعه مهم‌ترین عامل تعیین کننده وزن برآوردگرها می‌باشد [۴۳].

۳-۳- شبکه عصبی پرسپترون چندلایه

شبکه‌های عصبی مصنوعی یکی از روش‌های محاسباتی و الهام گرفته از سامانه عصبی مغز انسان هستند [۴۴]. از مشخصه‌های قابل توجه این نوع از شبکه‌ها قابلیت یادگیری آنها و توانایی تعمیم این یادگیری است و به دلیل داشتن این امکان یادگیری، درک الگوها را ممکن می‌سازند [۴۵]. این نوع شبکه‌ها جهت تشخیص ارتباطات غیر خطی مورد استفاده قرار می‌گیرند [۴۶]. مهم‌ترین مزیت شبکه‌های عصبی مصنوعی نسبت به روش‌های رگرسیون به منظور پیش‌بینی و مدل‌سازی یک الگو، عدم نیاز به یک مدل اولیه در ارتباط دادن بین داده‌های ورودی و خروجی می‌باشد. در واقع بر اساس روابط ذاتی میان داده‌ها، مدلی خطی یا غیر خطی بین متغیرهای مستقل و وابسته برقرار می‌شود [۴۴]. این نوع شبکه‌ها ابزار قدرتمندی هستند و در حوزه‌های مختلفی همچون؛ مدل‌سازی و پیش‌بینی بیماری‌ها [۶]، [۲۲]، طبقه‌بندی کاربری زمین [۴۷]، مدل‌سازی تغییر کاربری اراضی [۴۸]، طبقه‌بندی الگو [۴۹] و پیش‌بینی آب و هوا [۵۰] کاربرد دارند. همانطور که در شکل (۱) نشان داده شده است، شبکه‌های عصبی عموماً از سه بخش اصلی تشکیل شده‌اند: لایه ورودی، لایه‌های میانی (مخفی) و لایه خروجی [۵۱]. با توجه به این که شبکه عصبی پرسپترون چند لایه به عنوان یکی از انواع روش‌های شبکه عصبی مصنوعی، مقید به فرض‌های اولیه توزیع داده‌ها نمی‌باشد و یکی از پرکاربردترین نوع شبکه‌های عصبی بشمار می‌آید، به طوری که نسبت به روش‌های دیگر بیشتر برای مدل‌سازی یک رویداد همچون مدل‌سازی و پیش‌بینی بیماری کاربرد دارد. از اینرو در این مطالعه از شبکه عصبی پرسپترون چند لایه جهت مدل‌سازی بیماری لپتوسپیروز استفاده شده است. این نوع از شبکه عصبی شامل مجموعه‌ای از نورون‌هاست که در لایه‌های مختلفی پشت سر هم قرار گرفته‌اند. قانون یادگیری پرسپترون چندلایه، قاعده پس انتشار خطا نام دارد که به منظور تخمین پارامترهای مجهول شبکه مورد

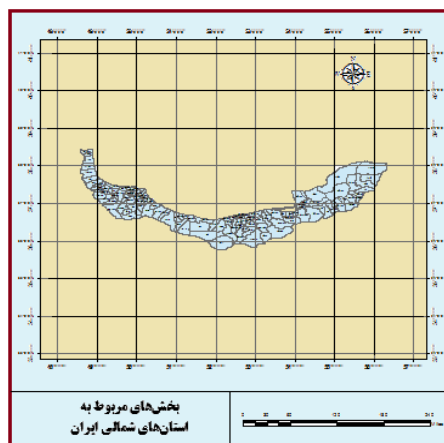
^۱ Root Mean Square Error

^۲ Mean Absolute Percentage Error

در نهایت ارزیابی آنها بوسیله معیارهای ضریب کاپا، RMSE، MAPE و R^2 تشریح شده‌اند.

۴-۱- محدوده تحقیق و داده‌های مورد استفاده

منطقه مورد مطالعه در این تحقیق استان‌های حاشیه دریای خزر به ترتیب گیلان، مازندران و گلستان در نظر گرفته شده‌اند. این استان‌ها از نظر تقسیمات سیاسی متشکل از ۱۱۹ بخش^۱ در مجموع می‌باشند که مربوط به زون‌های ۴۰ و ۴۱ شمالی سیستم تصویر UTM با طول جغرافیایی ۴۸ درجه و ۳۴ دقیقه تا ۵۶ درجه و ۱۹ دقیقه شرقی و عرض جغرافیایی ۳۵ درجه و ۴۶ دقیقه تا ۳۸ درجه و ۲۷ دقیقه شمالی هستند (شکل ۲). طبق اطلاعات اخذ شده حاصل از سرشماری سال ۱۳۹۰، جمعیت این استان‌ها به ۷,۳۳۱,۸۳۱ می‌رسد که ۳,۶۶۱,۴۹۸ افراد را مردان و ۳,۶۷۰,۳۳۳ افراد باقی‌مانده را زنان تشکیل می‌دهند. این استانها مساحتی حدود ۵۸,۲۵۰ کیلومتر مربع واقع در شمال ایران را می‌پوشانند. با توجه به موقعیت جغرافیایی و شرایط آب و هوایی، این استانها دارای آب و هوای معتدل و مرطوب می‌باشند.



شکل ۲- استانهای شمالی ایران به عنوان مناطق مورد مطالعه

مطابق با شواهد و مطالعات گذشته متغیرهای محیطی و شرایط آب و هوایی در بروز بیماری لپتوسپیروز موثر می‌باشند. بنابراین در این تحقیق به منظور تهیه نقشه پیش‌بینی بیماری لپتوسپیروز، داده‌های مربوط به وقوع بیماری، داده‌های جمعیتی و سرشماری، عوامل و متغیرهای محیطی و داده‌های توپوگرافی مورد استفاده قرار گرفتند.

داده‌ها از منابع مختلفی جمع‌آوری شده‌اند (جدول ۱). داده‌های مربوط به سرشماری و جمعیتی از مرکز آمار ایران بدست آمده‌اند. داده‌های مربوط به وقوع بیماری لپتوسپیروز توسط مرکز کنترل و پیشگیری بیماری‌های واگیر وزارت بهداشت، درمان و آموزش پزشکی ایران ثبت و ضبط شده‌اند که این داده‌ها شامل تعداد موارد بیماری، تاریخ بروز بیماری، سن، جنس بیمار و آدرس محل سکونت بیمار می‌باشند. سپس با استفاده از داده‌های جمعیتی و وقوع بیماری لپتوسپیروز نرخ شیوع استاندارد^۲ بیماری لپتوسپیروز مطابق با معادله (۶)، محاسبه گردید.

$$(۶) \quad ۱۰۰۰۰۰ * \frac{\text{تعداد افراد مبتلا در یک جمعیت در یک زمان مشخص}}{\text{تعداد کل جمعیت در همان زمان}} = \text{نرخ شیوع}$$

داده‌های توپوگرافی به منظور تولید نقشه‌های ارتفاعی، شیب و جهت شیب مورد استفاده قرار گرفتند. در واقع از مدل رقومی ارتفاعی (DEM)^۳ استخراج شده توسط تصاویر ماهواره‌ای SRTM با قدرت تفکیک ۳۰ متر به منظور تهیه داده‌های ارتفاعی در منطقه مورد مطالعه استفاده گردید. در نهایت، داده‌های محیطی از قبیل میانگین دمای هوا، مجموع بارش سالیانه، میانگین رطوبت هوا و تعداد روزهای یخبندان به صورت ماهانه از سازمان هواشناسی ایران جمع‌آوری شده‌اند.

جدول ۱- خلاصه‌ای از داده‌های مورد استفاده در تحقیق

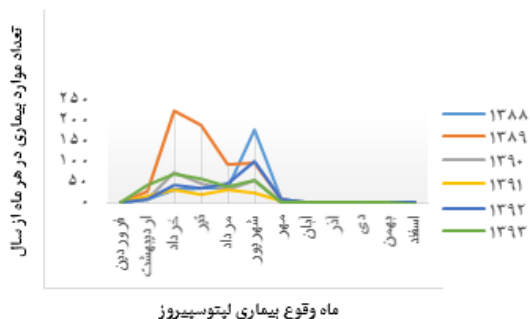
داده	توصیفات	
	منبع	نقشه مینا
توپوگرافی	نقشه DEM استخراج شده از تصاویر ماهواره‌ای SRTM با قدرت تفکیک ۳۰ متر	ارتفاع شیب جهت شیب
آب و هوایی	سازمان هواشناسی ایران	متوسط دما بارش متوسط رطوبت تعداد روزهای یخبندان
موارد بیماری لپتوسپیروز ثبت شده	مرکز کنترل و پیشگیری بیماری‌های واگیر وزارت بهداشت ایران	نرخ شیوع استاندارد (SIR)
جمعیتی و سرشماری	مرکز آمار ایران	نرخ شیوع استاندارد (SIR)

^۲ Standardized Incidence Rate (SIR)

^۳ Digital Elevation Model

^۱ District

طبق تاریخ‌های ثبت شده وقوع بیماری بیشترین موارد ابتلا به بیماری لپتوسپیروز در فاصله زمانی بین خرداد ماه تا پایان شهریور هر سال (ماه‌های فعالیت شالیکاران و افزایش دما) گزارش شده است (شکل ۵).

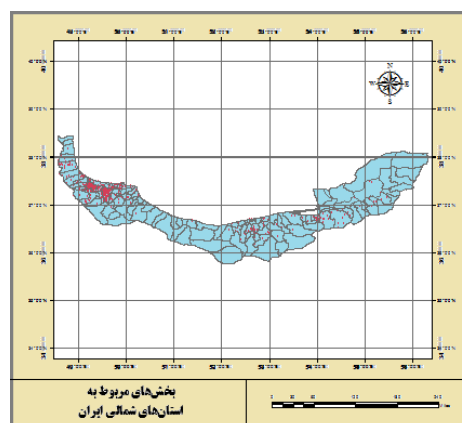


شکل ۵- فصول وقوع بیماری لپتوسپیروز (۱۳۸۸-۱۳۹۳)

۲-۴- بررسی وجود خودهمبستگی و خوشه‌ها

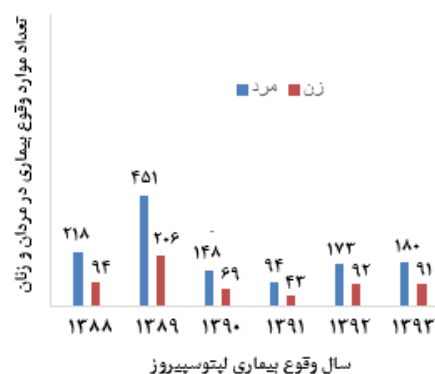
در این تحقیق تحلیلهای زمین آمار مورد استفاده متشکل از ۲ بخش اصلی می‌باشند. در ابتدا به منظور بررسی وجود یا عدم وجود خودهمبستگی مکانی میان موارد وقوع بیماری که آیا توزیع بیماری در منطقه تصادفی، یکنواخت یا خوشه‌ای است از شاخص‌های عمومی اندازه‌گیری خودهمبستگی مکانی همچون General G و شاخص Moran's I در سطح بخش‌های منطقه مورد مطالعه و به ترتیب برای هر ۶ سال مورد بررسی استفاده شده است. در گام بعدی جهت کشف خوشه‌های مکانی و شناسایی مناطقی که نرخ شیوع بیماری در آن مناطق با توجه به جمعیت تحت خطر به طور قابل توجهی از سایر مناطق بیشتر است از شاخص‌های محلی کشف خوشه‌های مکانی همچون شاخص 'Local Moran's I'، شاخص Local G_i استفاده شده است. با توجه به جدول (۲) آزمون‌های خودهمبستگی عمومی برای هر ۶ سال به طور جداگانه فرض صفر (توزیع بیماری در منطقه تصادفی است) را رد کرده و بیانگر این است که توزیع بیماری در منطقه در هر ۶ سال مورد بررسی خوشه‌ای می‌باشد. طبق جدول مقادیر p-value بدست آمده دارای مقادیر کوچکی می‌باشند و نشان می‌دهند که نتایج حاصل از آزمون‌های خودهمبستگی از لحاظ آماری معنی‌دار بوده و شرط درست بودن فرض صفر بر اساس داده‌های مشاهده شده رد می‌شود و توزیع بیماری در منطقه به صورت تصادفی نمی‌باشد. آماره p-value

در این تحقیق داده‌های مربوط به بیماری لپتوسپیروز در ۳ استان شمالی ایران از ابتدای فروردین ماه ۱۳۸۸ تا انتهای اسفند ماه ۱۳۹۳ به صورت ماهانه در سطح بخش‌های استانها، مورد استفاده قرار گرفتند. طی این سالها جمعا ۱۸۵۹ مورد بیماری لپتوسپیروز ثبت شده است. بررسی‌های آماری از داده‌های جمع‌آوری شده نشان می‌دهد که اکثریت موارد بیماری در ۶ سال مورد بررسی به ترتیب مربوط به استان‌های گیلان (۷۳ درصد)، مازندران (۱۸ درصد) و گلستان (۹ درصد) می‌باشد (شکل ۳).



شکل ۳- توزیع مکانی موارد بیماری لپتوسپیروز در سال‌های ۸۸-۹۳

در میان ۶ سال مورد مطالعه سال ۱۳۸۹ با ۶۵۷ مورد بیماری (۳۵ درصد) بالاترین موارد بیماری را به خود اختصاص داده است. موارد بیماری در هر دو جنسیت مردان و زنان مشاهده شده است. این بیماری در مردان ۱۲۶۴ مورد (۶۸ درصد) و در زنان ۵۹۵ مورد (۳۲ درصد) اتفاق افتاده است که به نظر می‌رسد به علت ویژگی‌های رفتاری مرد مانند پوشش کمتر و تعاملات اجتماعی و محیطی بیشتر، شیوع بیماری در مردان بیشتر می‌باشد (۵۴) (شکل ۴).



شکل ۴- پراکندگی موارد بیماری لپتوسپیروز در مردان و زنان

تقریبی از مساحت زیر منحنی توزیع نرمال است که با آزمون آماری محدود می‌شود و به منظور ارزیابی اینکه آیا نتایج حاصل از آزمون به دلیل تصادف و شانس رخ داده است یا خیر، استفاده می‌شود. این مقدار هر چه کوچکتر و

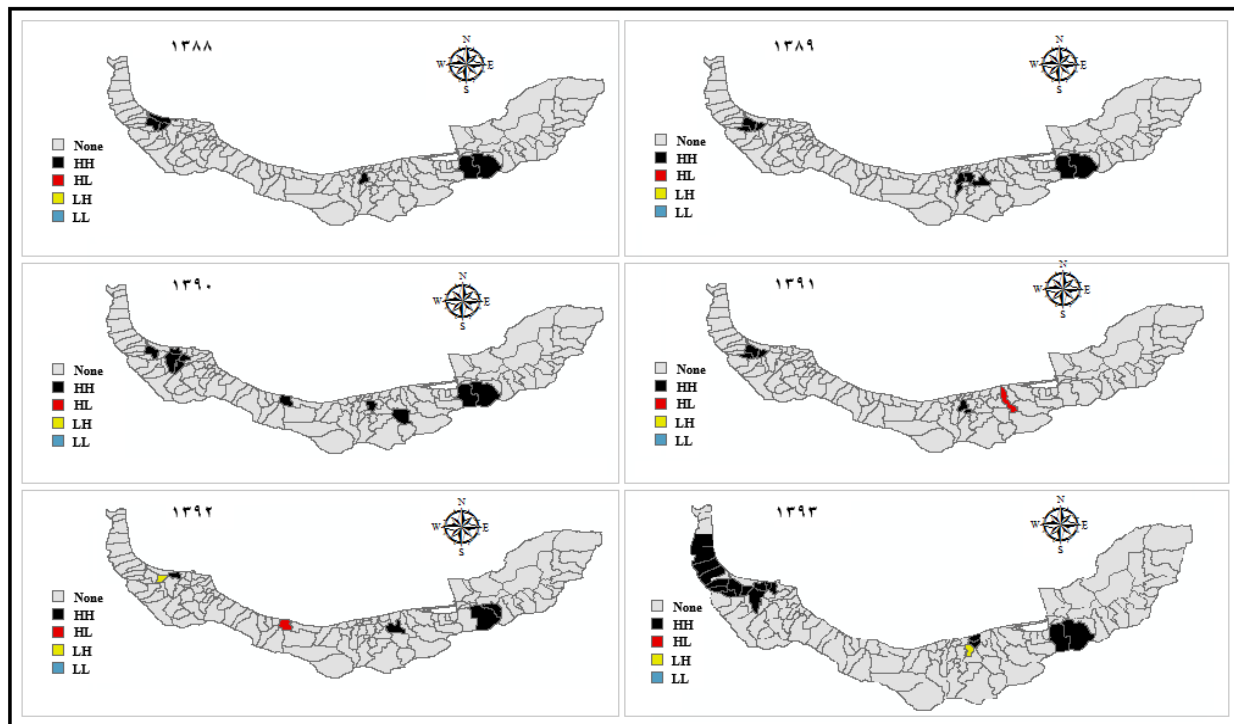
به صفر نزدیک‌تر باشد احتمال تصادفی و شانس بودن نتیجه حاصل از آزمون مورد بررسی کمتر و سطح معناداری بیشتر می‌شود [۳۹].

جدول ۲- نتایج آزمون‌های خودهمبستگی عمومی توزیع بیماری لپتوسپیروز (۱۳۸۸-۱۳۹۳).

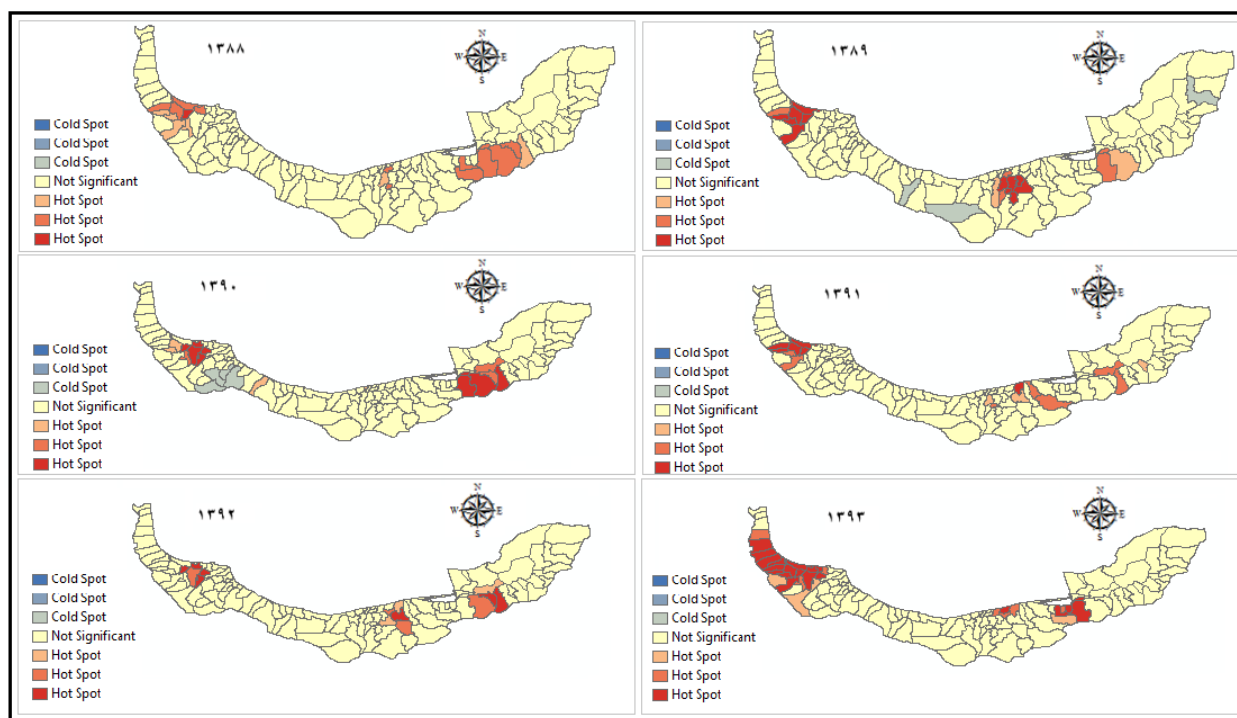
سال	آزمون عمومی	مقدار شاخص	Z – score	P – value	نوع توزیع
۱۳۸۸	شاخص General G	۰	۲/۹۶۹۹	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۲۲	۳/۲۶۷۴	<۰/۰۱	خوشه‌ای
۱۳۸۹	شاخص General G	۰	۳/۹۱۱۴	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۱۹	۳/۹	<۰/۰۱	خوشه‌ای
۱۳۹۰	شاخص General G	۰	۵/۶۴۲۵	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۴۵	۶/۴۷	<۰/۰۱	خوشه‌ای
۱۳۹۱	شاخص General G	۰	۴/۰۰۰۹	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۳۲	۵/۱۹	<۰/۰۱	خوشه‌ای
۱۳۹۲	شاخص General G	۰	۳/۹۴۱۱	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۲۵	۴/۴۴	<۰/۰۱	خوشه‌ای
۱۳۹۳	شاخص General G	۰	۳/۶۲۵۸	<۰/۰۱	خوشه‌ای
	شاخص Moran's I	۰/۲۸	۳/۸۷	<۰/۰۱	خوشه‌ای

خوشه‌هایی که توسط شاخص Local Moran's I شناسایی و کشف می‌شوند در چهار گروه قرار می‌گیرند. ناحیه بالا- بالا (HH) مکانی را نشان می‌دهد که در آن منطقه‌ای با مقدار بالای شاخص در همسایگی مناطقی قرار دارد که آن مناطق نیز دارای مقادیر بالای این شاخص می‌باشند. ۲) ناحیه بالا- پایین (HL) بیانگر مکانی است که در آن منطقه‌ای با مقدار بالای شاخص در مجاورت مناطقی با مقدار پایین این شاخص قرار دارد. ۳) ناحیه پایین- بالا (LH) شامل منطقه‌ای می‌باشد که مقدار شاخص در آن پایین است ولی در همسایگی آن مناطقی با مقدار شاخص بالا وجود دارند. ۴) ناحیه پایین- پایین (LL) نشان دهنده منطقه‌ای است که دارای مقدار پایین شاخص بوده و در هم‌جواری آن هم مناطقی با شاخص پایین وجود دارند. همچنین خوشه‌های استخراج شده در شاخص Local G_i شامل شش گروه از قبیل مناطق با ریسک بالا و پرخطر (Hot Spot) در سطوح اطمینان ۹۹درصد، ۹۵ درصد و ۹۰ درصد و مناطق با ریسک پایین و کم‌خطر (Cold Spot) در سطوح اطمینان ۹۹درصد، ۹۵ درصد و ۹۰ درصد می‌باشد. از طرفی مناطقی که شامل

هیچ‌یک از موارد فوق نباشند، به عنوان مناطق بدون اهمیت از حیث مکانی شناخته می‌شوند. نواحی HH و Hot Spot، مناطقی را نشان می‌دهند که خوشه‌های وجود بیماری در آن مناطق حضور دارند. همچنین نواحی LL و Cold Spot، بیانگر مناطقی هستند که خوشه‌های عدم وجود بیماری در آن مناطق حضور دارند [۳۹]. همانطور که در شکل‌های (۶) و (۷) نشان داده شده است، نتایج حاصل از آزمون‌های کشف خوشه‌های مکانی بیماری لپتوسپیروز در منطقه مورد مطالعه با هم مطابقت دارند. این نتایج نشان می‌دهند که خوشه‌ها تقریباً در مرکز و غرب استان گیلان، نواحی شمالی و شمال شرقی استان مازندران و مناطق غربی استان گلستان اتفاق افتاده‌اند. از دلایل چنین نتایجی می‌توان به وجود زمین‌های کشاورزی و شالیکاری بیشتر نسبت به سایر مناطق، بارش فراوان باران و همچنین ارتفاع و شیب پائین این مناطق و در نتیجه افزایش تجمع آب‌های سطحی و راکد اشاره نمود که باعث افزایش احتمال بروز بیماری لپتوسپیروز در این مناطق می‌باشد.



شکل ۶- کشف خوشه‌های مکانی با استفاده از شاخص Moran's I



شکل ۷- کشف خوشه‌های مکانی با استفاده از شاخص General G

۴-۳- مدل‌سازی بیماری

به منظور تهیه نقشه پیش‌بینی بیماری لپتوسپیروز به روش رگرسیون وزن‌دار جغرافیایی، از داده‌های محیطی و توپوگرافی موثر مربوط به سال ۱۳۹۲ به همراه نرخ شیوع

بیماری لپتوسپیروز سال ۱۳۹۱ برای بدست آوردن نکویی برازش^۱ (R^2) روش رگرسیون استفاده شده است (جدول ۳). سپس از داده‌های محیطی و توپوگرافی موثر مربوط به

^۱ Goodness of fit

سال ۱۳۹۳ به همراه نرخ شیوع بیماری لپتوسپیروز سال ۱۳۹۲ جهت تهیه نقشه پیش‌بینی نرخ شیوع بیماری سال ۱۳۹۳ استفاده گردیده است (جدول ۴). با توجه به اینکه بیماری لپتوسپیروز از جمله بیماری‌های واگیر محسوب می‌شود، لذا در این تحقیق نرخ شیوع بیماری در سال‌های گذشته نیز به عنوان یکی از متغیرهای مستقل در مدل‌سازی در نظر گرفته شده است تا مشخص گردد که آیا تعداد موارد وقوع بیماری در سال‌های گذشته تاثیری در تعداد موارد وقوع بیماری در سال جاری دارد یا خیر و اینکه آیا امکان انتقال بیماری از انسان به انسان امکان‌پذیر می‌باشد یا خیر. به منظور به‌کارگیری هر کدام از متغیرها در مدل‌سازی ابتدا لایه‌های اطلاعاتی مورد نیاز اعم از نقشه‌های ارتفاعی، شیب، جهت شیب، میانگین دمای هوا، میانگین رطوبت هوا، مجموع بارش سالیانه، تعداد روزهای یخبندان و نرخ شیوع بیماری لپتوسپیروز به فرمت شیب فایل در نرم افزار ArcGIS 10.2.2 ایجاد شدند. با توجه به اینکه داده‌های هواشناسی به صورت نقطه‌ای وجود دارند، از روش درونیابی IDW برای داشتن سطحی پیوسته استفاده شده است. بدین ترتیب که با استفاده از روش IDW هر لایه نقطه‌ای به یک لایه رستری تبدیل شده و لایه رستری مربوطه براساس محدوده منطقه مورد مطالعه برش داده شده است. سپس برای هر پلی‌گون متوسط پیکسل‌های داخل پلی‌گون، به عنوان پارامتر آن پلی‌گون در نظر گرفته شد. در نهایت برای یکسان نمودن دامنه مقادیر متغیرها و به حداقل رساندن اثر مقیاس، مقادیر متغیرها در بازه [۰-۱] نرمال‌سازی شده‌اند. به منظور آگاهی از هم خطی بودن متغیرهای مستقل از شاخص VIF^1 استفاده شده است. VIF شاخصی است که برای آگاهی از وجود یا عدم وجود رابطه هم‌خطی بین متغیرهای مستقل استفاده می‌شود. این شاخص شدت هم‌خطی میان متغیرهای مستقل (هم‌خطی چندگانه) را نشان می‌دهد. در واقع این شاخص بیان می‌کند که چه مقدار از تغییرات مربوط به ضرایب برآورد شده برای هر متغیر مستقل بابت برقراری رابطه هم‌خطی بین متغیرهای مستقل افزایش یافته است. کمترین مقدار این شاخص مثبت یک و بیشترین مقدار آن مثبت بی‌نهایت می‌باشد. به عنوان یک قاعده تجربی، اگر مقدار VIF بزرگتر از ۷/۵

باشد، نشان‌دهنده هم‌خطی چندگانه بالا می‌باشد (در برخی موارد اعداد ۵ یا ۱۰ نیز به عنوان آستانه معرفی می‌گردند) [۵۵]. بر اساس نتایج بدست آمده در جدول (۳) شاخص VIF برای متغیرهای مورد استفاده به قدری نیست که نتوان آنها را در مدل‌سازی دخالت داد ($VIF < 1$ 7.5). بنابراین فرض مستقل بودن متغیرهای ورودی فرض درستی بوده است و می‌توان از تمامی این متغیرها در مدل‌سازی بیماری استفاده نمود.

جدول ۳- نتایج بدست آمده از رگرسیون وزن‌دار جغرافیایی

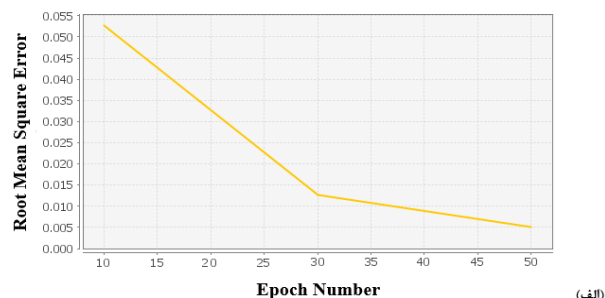
متغیر وابسته	متغیر مستقل مربوط به سال ۱۳۹۲	VIF	R ²
نرخ شیوع بیماری ۱۳۹۲	ارتفاع	۱/۳۹۷۴	۰/۳۲
	شیب	۱/۴۳۲۶	۰/۲۴
	جهت شیب	۱/۵۸۴۹	۰/۰۷
	دما	۱/۰۲۲۱	۰/۳۱
	رطوبت	۱/۲۵۴۸	۰/۴۲
	بارش	۱/۱۷۵۹	۰/۳۷
	تعداد روزهای یخبندان	۱/۴۱۱۷	۰/۱۲
	نرخ شیوع بیماری ۱۳۹۱	۱/۶۳۲۵	۰/۰۹

جدول ۴- ضریب تعیین بدست آمده از مدل‌سازی به روش رگرسیون

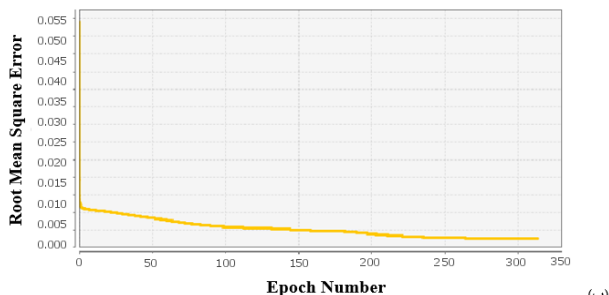
متغیر وابسته	متغیر مستقل مربوط به سال ۱۳۹۳	VIF	R ²
نرخ شیوع بیماری ۱۳۹۳	ارتفاع	۱/۱۸۹۶	نکویی برازش حاصل از مدل = ۰/۷۶
	شیب	۱/۳۹۶۰	
	جهت شیب	۲/۰۱۳۲	
	دما	۱/۱۷۱۹	
	رطوبت	۱/۵۰۹۱	
	بارش	۱/۰۶۰۲	
	تعداد روزهای یخبندان	۱/۶۲۸۶	
	نرخ شیوع بیماری ۱۳۹۲	۱/۸۹۳۴	

به منظور آموزش شبکه عصبی MLP جهت تهیه نقشه پیش‌بینی مکانی-زمانی بیماری لپتوسپیروز به دو صورت عمل شده است. در حالت اول برای برقراری شرایط کاملاً یکسان با روش رگرسیون وزن‌دار جغرافیایی، از داده‌های مربوط به سال ۱۳۹۲ به همراه نرخ شیوع بیماری سال

^۱ Variance Inflation Factor



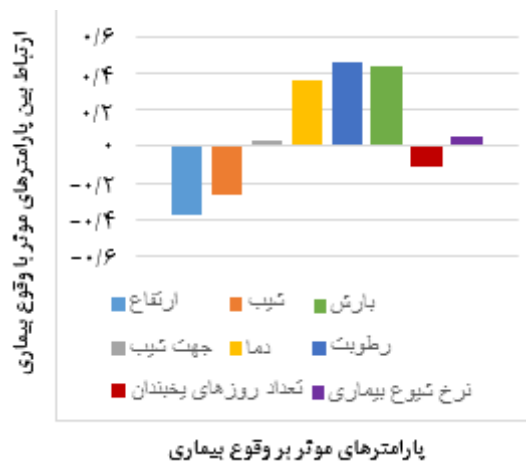
(الف)



(ب)

شکل ۸- نحوه آموزش شبکه عصبی MLP با داده‌های آموزشی. (الف) شبکه عصبی MLP در حالت اول. (ب) شبکه عصبی MLP در حالت دوم

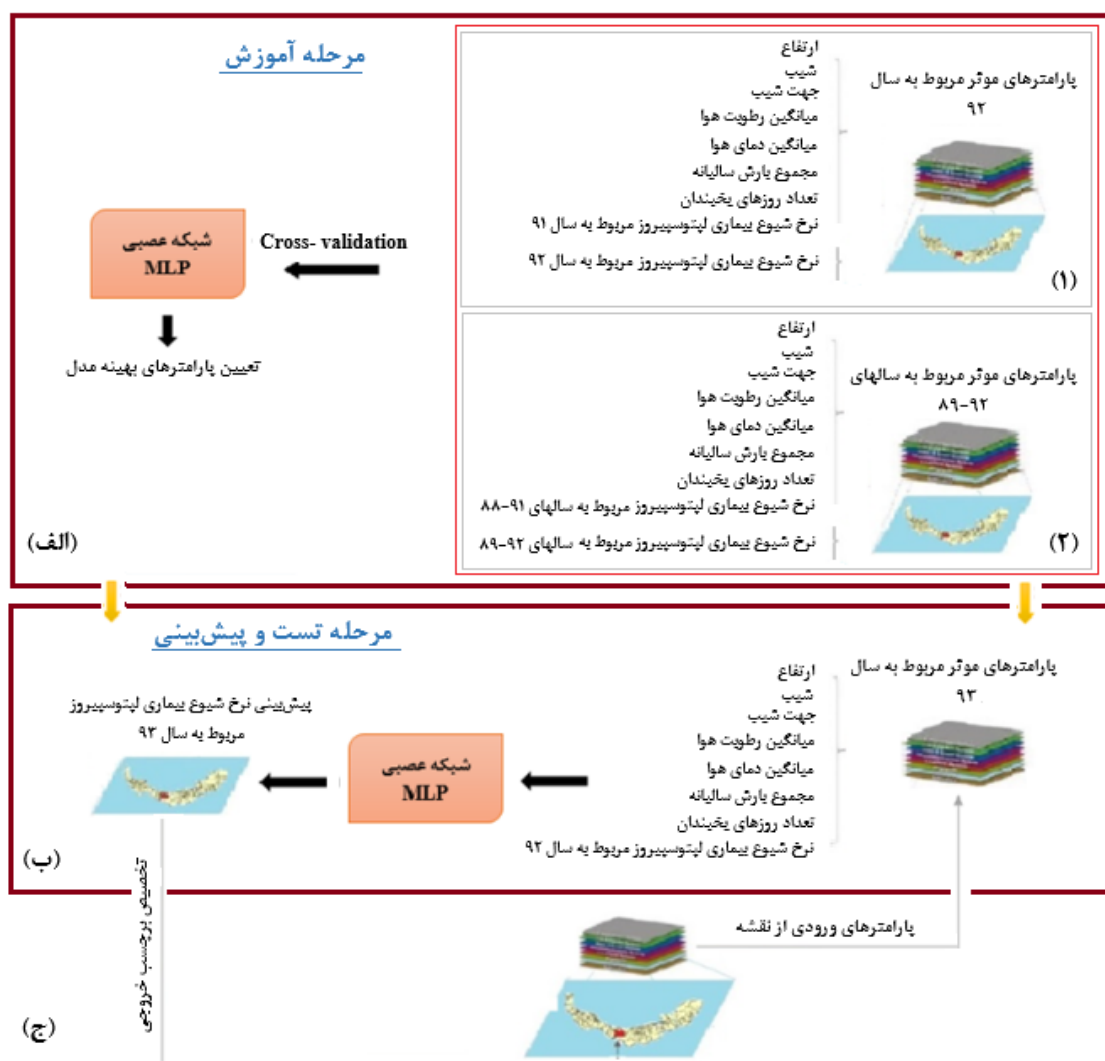
به منظور درک ارتباط و همبستگی میان متغیرهای موثر بر شیوع بیماری لپتوسپیروز با وقوع بیماری و تعیین متغیرهایی که تاثیر بیشتری بر روی وقوع بیماری دارند، از تحلیل داده‌ای توسط نرم افزار مایکروسافت اکسل ۲۰۱۳ استفاده شده است. همانطور که شکل (۱۰) نشان می‌دهد پارامترهای متوسط رطوبت، بارش سالیانه، متوسط دما، نرخ شیوع بیماری و جهت شیب، رابطه‌ای مثبت با وقوع بیماری لپتوسپیروز دارند و پارامترهای ارتفاع، شیب و تعداد روزهای یخبندان دارای ارتباط منفی با وقوع بیماری می‌باشند. همچنین می‌توان گفت که پارامترهای متوسط رطوبت، بارش، ارتفاع و متوسط دما صرف‌نظر از علامت‌های مثبت و منفی، به ترتیب ارتباط و همبستگی قوی‌تری با وقوع بیماری لپتوسپیروز دارا می‌باشند.



شکل ۱۰- ارتباط و همبستگی پارامترهای موثر با وقوع بیماری لپتوسپیروز

۱۳۹۱ با هدف آموزش شبکه عصبی استفاده شده است (شکل ۹ الف). همچنین به منظور ارزیابی مدل آموزش دیده و تهیه نقشه پیش‌بینی بیماری لپتوسپیروز سال ۱۳۹۳، داده‌های موثر مربوط به سال ۱۳۹۳ و نرخ شیوع بیماری سال ۱۳۹۲ وارد مدل شدند و دقت شبکه مورد ارزیابی قرار گرفت ($RMSE = 0.005$) (شکل ۸ الف، ۹ ب). در حالت دوم، پارامترهای موثر ورودی به همراه نرخ شیوع بیماری لپتوسپیروز مربوط به سال‌های ۱۳۸۸ تا ۱۳۹۲، به عنوان داده‌های آموزشی به شبکه عصبی MLP تغذیه شدند (شکل ۹ الف). سپس شبکه عصبی آموزش دیده با هدف تهیه نقشه پیش‌بینی بیماری لپتوسپیروز در سال ۱۳۹۳ استفاده شد و دقت شبکه مورد ارزیابی قرار گرفت ($RMSE = 0.0028$) (شکل ۸ ب، ۹ ب). در هر دو حالت مدل شبکه عصبی MLP، به منظور تعیین پارامترهای بهینه مدل، تکنیک اعتبارسنجی متقابل^۱ بر روی ۲۵ درصد از داده‌های آموزشی در مرحله آموزش شبکه، مورد استفاده قرار گرفت. دلیل استفاده از روش اعتبارسنجی متقابل، تعیین مقادیر بهینه پارامترهای مدل برای کاربرد مورد نظر می‌باشد به گونه‌ای که مدل بتواند داده‌های نامعلوم (داده‌های ارزیابی) را با دقت خوبی پیش‌بینی کند. در این روش داده‌های آموزشی به دو قسمت تقسیم می‌شوند، یک قسمت به منظور آموزش مدل استفاده می‌شوند و قسمت دیگر برای ارزیابی مدل مورد استفاده قرار می‌گیرند. درصد صحت اندازه‌گیری شده در این قسمت می‌تواند معیار خوبی برای ارزیابی عملکرد مدل هنگام پیش‌بینی داده‌های نامعلوم باشد. نحوه عملکرد این روش بدین صورت است که ابتدا داده‌های آموزشی به M زیر مجموعه مساوی تقسیم می‌شوند. $M-1$ زیرمجموعه برای آموزش استفاده می‌شوند و یک مجموعه برای ارزیابی عملکرد مدل مورد استفاده قرار می‌گیرد. این عمل M بار تکرار می‌شود و در هر بار یک زیرمجموعه به عنوان زیرمجموعه ارزیابی در نظر گرفته می‌شود. به این ترتیب تمام داده‌های آموزشی یک بار برای ارزیابی عملکرد مدل استفاده می‌شوند. بنابراین با استفاده از این روش مقادیر بهینه پارامترهای مدل به منظور برقراری تعادل در قدرت تمایز و قدرت تعمیم‌دهی مدل انتخاب می‌شوند [۵۶]. در نهایت مدل شبکه عصبی MLP آموزش دیده قادر است نرخ شیوع بیماری لپتوسپیروز را در سطح بخش‌های استان‌های مورد مطالعه برای سال‌های آینده پیش‌بینی کند (شکل ۹ ج).

^۱ Cross- validation



شکل ۹- (الف) فرآیند آموزش شبکه عصبی MLP. (ب) فرآیند ارزیابی دقت پیش‌بینی شبکه عصبی MLP.

نشان می‌دهد که چند درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل قابل برآورد می‌باشد [۵۷]. این معیار ارزیابی با استفاده از رابطه (۷) قابل محاسبه می‌باشد:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (7)$$

که در این رابطه: N ، تعداد پیش‌بینی‌ها، y_i ، $\hat{y}_i$ و $\bar{y}$ به ترتیب بیانگر مقادیر واقعی، پیش‌بینی شده و میانگین مقادیر واقعی می‌باشند.

در معیار ارزیابی ماتریس مقایسه، با استفاده از ضریب کاپا میزان سازگاری مدل پیش‌بینی شده و واقعیت بررسی شده است. ضریب کاپا روش آماری است که در سطح وسیعی در مسائل مکانی بکار گرفته می‌شود و میزان سازگاری مدل پیش‌بینی شده و واقعیت را نشان می‌دهد [۵۸]. با استفاده از این ضریب می‌توان میزان شباهت دو

۴-۴- ارزیابی نتایج حاصل از مدلها و تعیین اهمیت متغیرهای مورد استفاده در مدلسازی

به منظور ارزیابی نقشه‌های پیش‌بینی نرخ شیوع بیماری لپتوسپیروز در سال ۱۳۹۳ حاصل از دو روش شبکه عصبی پرسپترون چند لایه و رگرسیون وزندار جغرافیایی علاوه بر معیارهای RMSE و MAPE، از معیارهای R^2 و ماتریس مقایسه استفاده شده است. معیار ارزیابی R^2 در علم آمار بیانگر میزان خوب بودن مدل آماری در تخمین متغیر وابسته بر اساس متغیرهای مستقل می‌باشد که مقداری بین ۰ تا ۱ دارد. مقدار ۱ بیانگر این موضوع است که مدل استفاده شده قادر است به طور کامل متغیر وابسته را با در نظر گرفتن متغیرهای مستقل پیش‌بینی کند و مقدار ۰ نشان‌دهنده آن است که نمی‌توان با متغیرهای مستقل در نظر گرفته شده متغیر وابسته را تخمین زد. به عبار دیگر، این معیار ارزیابی

$$Kappa = \frac{P(O) - P(E)}{1 - P(E)}$$

$$P(O) = \sum_{i=1}^C P_{ii} \quad (A)$$

$$P(E) = \sum_{i=1}^C P_{i.} * P_{.i}$$

که در رابطه فوق داریم :

$P(O)$ و $P(E)$ به ترتیب بیانگر مجموع نسبت تعداد موارد دارای ارزش (کلاس) مشابه و متفاوت در دو نقشه واقعی و پیش‌بینی شده برای ارزش‌های متفاوت می‌باشند.

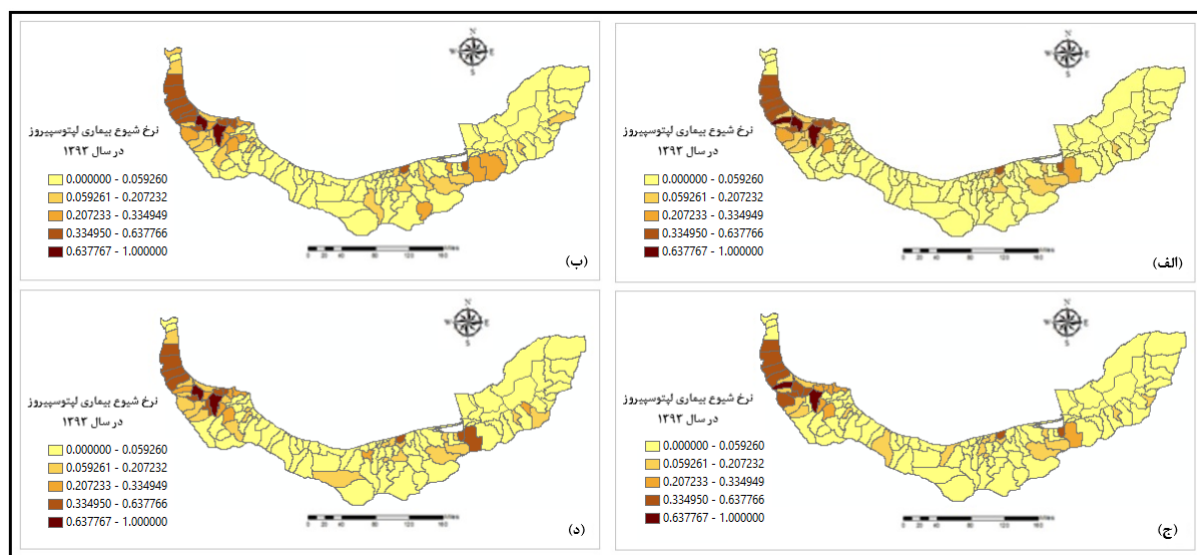
P_{ii} ، $P_{i.}$ و $P_{.i}$ به ترتیب بیانگر تعداد مواردی که در دو نقشه دارای ارزش i می‌باشند، تعداد مواردی که در نقشه واقعی دارای ارزش i و در نقشه پیش‌بینی دارای ارزش غیر i می‌باشند، تعداد مواردی که در نقشه پیش‌بینی دارای ارزش i و در نقشه واقعی دارای ارزش غیر i می‌باشند.

نقشه واقعیت و پیش‌بینی نرخ شیوع بیماری لپتوسپیروز حاصل از مدل‌های رگرسیون و زنادار جغرافیایی و شبکه عصبی پرسپترون چند لایه در شکل (۱۱) و نتایج ارزیابی دو مدل در جدول (۶) نشان داده شده‌اند.

نقشه را در پخش مکانی اندازه گرفت [۵۹]. برای محاسبه ضریب کاپا از جدول احتمال وقوع (جدول ۵) استفاده می‌شود که عناصر آن از ماتریس مقایسه بدست می‌آیند. در واقع ضریب کاپا از کلیه عناصر ماتریس مقایسه جهت بررسی سازگاری لایه پیش‌بینی شده و واقعیت استفاده می‌کند، بنابراین روش دقیقی محسوب می‌گردد [۵۹]. ضریب کاپا با استفاده از رابطه (۸) محاسبه می‌شود [۵۹].

جدول ۵- جدول احتمال وقوع

واقعیت	مدل					جمع
	کلاس	1	2	...	C	
	1	P_{11}	P_{12}	...	P_{1C}	P_{1T}
	2	P_{21}	P_{22}	...	P_{2C}	P_{2T}
	...	...	...	...	...	...
	C	P_{C1}	P_{C2}	...	P_{CC}	P_{CT}
جمع		P_{T1}	P_{T2}	...	P_{TC}	1



شکل ۱۱- (الف) نقشه واقعیت نرخ شیوع بیماری لپتوسپیروز. (ب) نقشه پیش‌بینی نرخ شیوع بیماری لپتوسپیروز با استفاده از رگرسیون و زنادار جغرافیایی. (ج) نقشه پیش‌بینی نرخ شیوع بیماری لپتوسپیروز با استفاده از شبکه عصبی پرسپترون چندلایه در حالت اول. (د) نقشه پیش‌بینی نرخ شیوع بیماری لپتوسپیروز با استفاده از شبکه عصبی پرسپترون چندلایه در حالت دوم.

جدول ۶- نتایج ارزیابی دو مدل

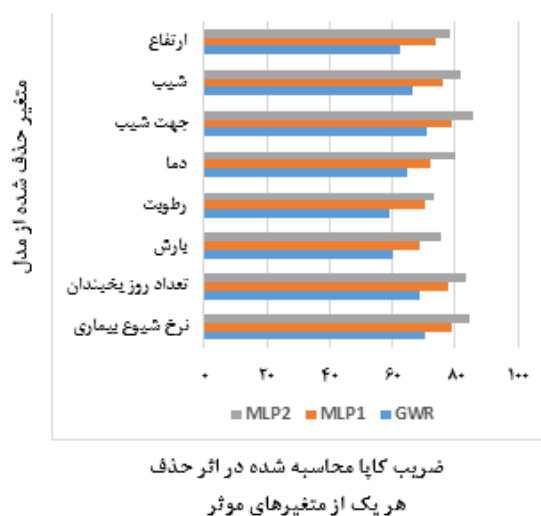
روش	تعداد موارد درست پیش‌بینی شده	تعداد موارد اشتباه پیش‌بینی شده	درصد تناظر صحیح	ضریب کاپا (درصد)	MAPE	RMSE	R ²
شبکه عصبی MLP (حالت دوم)	۱۰۷	۱۲	۸۹/۹۱	۸۷/۳۸	۱/۰۶	۰/۰۰۲۸	۰/۹۰
شبکه عصبی MLP (حالت اول)	۱۰۵	۱۴	۸۸/۲۳	۸۱/۵۸	۱/۱۷	۰/۰۰۵	۰/۸۵
رگرسیون (GWR)	۱۰۳	۱۶	۸۶/۵۵	۷۳/۴۴	۱/۳۶	۰/۰۱۳	۰/۷۶

ضرایب بدست آمده برای متغیرهای موثر ورودی روش رگرسیون وزندار جغرافیایی در جدول (۷) نشان داده شده است. ضرایب بدست آمده برای هر بخش متفاوت می‌باشد، لذا مقادیر حداقل، حداکثر و میانگین حاصل‌شده در این

جدول قرار داده شده است. میانگین محاسبه شده از میانگین‌گیری ضرایب همه بخش‌ها بدست آمده و انحراف معیار نیز میانگین همه انحراف معیارهای بخش‌هاست.

جدول ۷- ضرایب حاصل از مدل رگرسیون وزندار جغرافیایی

انحراف معیار	میانگین	حداکثر	حداقل	متغیر موثر ورودی
۵/۶۲	-۱/۰۵	۲۹/۳۶	-۲۲/۴۲	ارتفاع
۶/۸۵	-۰/۵۱	۲۰/۴۶	-۱۸/۷۲	شیب
۱/۱۲	۰/۰۴	۱/۱۴	-۵/۸۴	جهت شیب
۲/۱۶	۰/۹۵	۱۲/۵۹	-۴/۲۷	دما
۳/۵۹	۲/۵۳	۱۲/۶۸	-۵/۳۴	رطوبت
۲/۴۴	۱/۲۷	۱۰/۱۴	-۸/۲۲	بارش
۵/۱۷	-۰/۱۹	۱۴/۷۶	-۸/۵۳	تعداد روزهای یخبندان
۱/۹۲	۰/۰۸	۱/۰۸	-۶/۲۶	نرخ شیوع بیماری ۱۳۹۲



شکل ۱۲- اولویت‌دهی متغیرهای موثر در مدلسازی

از آنجاییکه در روش شبکه عصبی بر خلاف روش رگرسیون وزندار جغرافیایی، از نحوه وزن‌دهی و ضرایب استفاده شده در آن اطلاعی نداریم، لذا به منظور تعیین اهمیت و اولویت‌دهی به هر متغیر پیش‌بینی کننده و تعیین میزان تاثیر هر یک از متغیرها در مدلسازی از آنالیز حساسیت استفاده شده است. آنالیز حساسیت بدین صورت عمل شده است که هر بار یک متغیر از بین متغیرهای موثر بر شیوع بیماری و استفاده شده در مدلسازی، از مدل خارج شده و مدل با استفاده از سایر متغیرها اجرا می‌شود و سپس عملکرد مدل با حذف آن متغیر سنجیده می‌شود. متغیری که کارایی مدل را بیشتر از سایر متغیرها کاهش می‌دهد و بیشترین تاثیر را بر روی مدل دارد به عنوان مهم‌ترین متغیر انتخاب می‌شود. نتایج حاصل از تعیین اهمیت متغیرها در شکل (۱۲) و جدول (۸) نشان داده شده است. لازم به ذکر است که آنالیز حساسیت برای روش رگرسیون وزندار جغرافیایی نیز اجرا شده است. نتایج حاصل از آنالیز حساسیت روش رگرسیون وزندار جغرافیایی با نتایج بدست آمده از جدول (۷) در رابطه با تعیین اهمیت و میزان تاثیر هر یک از متغیرها در مدلسازی به روش GWR مطابقت دارد.

جدول ۸- نتایج حاصل از تعیین اهمیت و میزان تاثیر متغیرهای موثر در مدل سازی

متغیر حذف شده از مدل	مدل رگرسیون وزندار جغرافیایی (GWR)				مدل شبکه عصبی در حالت اول (MLP1)				مدل شبکه عصبی در حالت دوم (MLP2)			
	ضریب کاپا (درصد)	MAPE	RMSE	R ²	ضریب کاپا (درصد)	MAPE	RMSE	R ²	ضریب کاپا (درصد)	MAPE	RMSE	R ²
ارتفاع	۶۲/۴۵	۱/۴۸	۰/۰۲۲۹	۰/۶۹	۷۳/۵۸	۱/۳۳	۰/۰۱۲۶	۰/۷۸	۷۸/۲۱	۱/۲۷	۰/۰۱۰۸	۰/۸۰
شیب	۶۶/۳۹	۱/۴۴	۰/۰۱۸۹	۰/۷۲	۷۵/۹۶	۱/۲۹	۰/۰۱۱۲	۰/۸۰	۸۱/۷۹	۱/۲۰	۰/۰۰۷۳	۰/۸۳
جهت شیب	۷۱/۲۹	۱/۳۹	۰/۰۱۴۱	۰/۷۵	۷۸/۸۲	۱/۲۲	۰/۰۰۸۳	۰/۸۳	۸۵/۵۳	۱/۰۹	۰/۰۰۳۴	۰/۸۸
دما	۶۴/۵۸	۱/۴۵	۰/۰۲۰۷	۰/۷۰	۷۲/۱۸	۱/۳۵	۰/۰۱۴۲	۰/۷۷	۸۰/۰۶	۱/۲۴	۰/۰۰۸۸	۰/۸۱
رطوبت	۵۹/۱۳	۱/۵۲	۰/۰۲۴۳	۰/۶۶	۷۰/۴۵	۱/۳۷	۰/۰۱۵۵	۰/۷۵	۷۳/۱۳	۱/۳۱	۰/۰۱۲۶	۰/۷۷
بارش	۶۰/۲۴	۱/۵۱	۰/۰۲۶۴	۰/۶۷	۶۸/۶۲	۱/۴۰	۰/۰۱۷۴	۰/۷۴	۷۵/۵۶	۱/۳۵	۰/۰۱۳۱	۰/۷۸
تعداد روزهای یخبندان	۶۸/۸۴	۱/۴۲	۰/۰۱۷۳	۰/۷۳	۷۷/۷۶	۱/۲۵	۰/۰۰۹۷	۰/۸۲	۸۳/۲۷	۱/۱۵	۰/۰۰۵۷	۰/۸۵
نرخ شیوع بیماری ۱۳۹۲	۷۰/۵۰	۱/۴۰	۰/۰۱۵۸	۰/۷۴	۷۹/۰۴	۱/۱۹	۰/۰۰۶۲	۰/۸۴	۸۴/۶۸	۱/۱۳	۰/۰۰۴۹	۰/۸۶

۵- جمع بندی و پیشنهادات

در این تحقیق، به منظور تهیه نقشه پیش بینی توزیع مکانی- زمانی بیماری لپتوسپیروز در استان های شمالی ایران، از روش های شبکه عصبی پرسپترون چند لایه و رگرسیون وزندار جغرافیایی، استفاده شده است. پارامترهای موثر در این مدل سازی، عبارتند از: میانگین دمای هوا، میانگین رطوبت، مجموع بارش سالیانه، ارتفاع، شیب، جهت شیب و تعداد روزهای یخبندان. بر طبق نتایج بدست آمده از رگرسیون، بین پارامترهای ورودی، وابستگی قابل ملاحظه ای مشاهده نشده است و بنابراین فرض مستقل بودن پارامترهای ورودی، فرض درستی بوده است. براساس تجزیه و تحلیل های آماری صورت گرفته، با صرف نظر از علامت های مثبت و منفی، پارامترهای متوسط رطوبت، بارش سالیانه، ارتفاع، متوسط دما و شیب به ترتیب موثرترین عوامل در بروز بیماری لپتوسپیروز می باشند و از طرف دیگر، تعداد روزهای یخبندان، نرخ شیوع بیماری و جهت شیب به ترتیب کم تاثیرترین عوامل در بروز بیماری لپتوسپیروز بشمار می آیند. همچنین در بررسی ارتباط و همبستگی متغیرها با وقوع بیماری، علامت ضرایب بدست آمده برای پارامترهای ارتفاع، شیب و تعداد روزهای یخبندان نشان می دهد که یک رابطه غیرمستقیم میان این عوامل و بروز بیماری لپتوسپیروز برقرار است به طوری که با افزایش ارتفاع، شیب و تعداد

روزهای یخبندان، از میزان وقوع بیماری لپتوسپیروز کاسته می شود (شکل ۱۰). تحلیل های زمین آمار بیماری لپتوسپیروز در هر ۶ سال خودهمبستگی مکانی معنی داری را نشان می دهند و بیانگر این امر هستند که در هر ۶ سال مورد مطالعه توزیع بیماری الگوی خوشه ای دارد. آزمون های کشف خوشه های مکانی نشان دادند که خوشه ها با بیشترین احتمال وقوع در هر ۶ سال در مرکز و غرب استان گیلان، نواحی شمالی و شمال شرقی استان مازندران و مناطق غربی استان گلستان به دلیل برخورداری از شرایط محیطی و آب و هوایی مناسب برای رشد و تکثیر ناقلین و مخازن بیماری رخ داده اند. همچنین بررسی های آماری حاکی از آن بوده اند که بیشترین موارد ابتلا به بیماری لپتوسپیروز در فاصله زمانی بین خرداد ماه تا پایان شهریور هر سال همزمان با شروع فصل کشاورزی و افزایش دما اتفاق افتاده است. مقادیر معیارهای ضریب کاپا، RMSE، MAPE و R² بدست آمده برای هر دو مدل پیشنهادی نشان می دهند که شبکه عصبی MLP در حالت دوم توانسته است نرخ شیوع بیماری لپتوسپیروز را بهتر از مدل رگرسیون وزندار جغرافیایی و حالت اول آموزش شبکه عصبی پیش بینی کند. علت این نتیجه را می توان بیان نمود که در حالت دوم به دلیل اینکه از داده های مربوط به سالهای بیشتری استفاده شده است شبکه قادر بوده تا با دقت بیشتری آموزش ببیند و در نتیجه خروجی دقیق تری حاصل نماید.

از نتایج حاصل از مدل‌سازی مشخص است که متوسط رطوبت و بارش سالیانه نقش مهمی را در تهیه نقشه پیش‌بینی توزیع مکانی- زمانی بیماری لپتوسپیروز دارا می‌باشند و مهمترین پارامترها از بین پارامترهای مورد استفاده در مدل‌سازی بیماری لپتوسپیروز در استان‌های شمالی ایران می‌باشند. همچنین با توجه به تاثیر اندک متغیر نرخ شیوع بیماری سال‌های گذشته در مدل‌سازی می‌توان به این نکته پی برد که اگرچه انتقال بیماری لپتوسپیروز از انسان به انسان بسیار به ندرت اتفاق می‌افتد، ولی نباید خطر ابتلا به این بیماری توسط افراد بیمار مورد غفلت قرار گیرد. در نهایت با استفاده از چنین نقشه‌هایی می‌توان مناطق آسیب‌پذیر و در معرض خطر بیماری لپتوسپیروز را شناسایی کرده و در راستای بهبود فرآیند تصمیم‌گیری بهینه در زمینه بهداشت عمومی در اختیار کارشناسان و مدیران قرار گیرد [۲۲]. به عنوان کارهای آینده، قصد داریم که اثر پارامترهای محیطی و اجتماعی- اقتصادی دیگر همچون پوشش و کاربری زمین، جنس خاک، سرعت وزش باد و غیره، به منظور بهبود و افزایش دقت مدل‌ها مورد ارزیابی قرار دهیم. همچنین در این مطالعه با توجه به محدودیت زمانی و داده‌ای،

مدلسازی بیماری به صورت سالیانه صورت پذیرفته است. لذا در مطالعات آینده با توجه به اینکه بیماری لپتوسپیروز در ماه‌ها و زمان‌های خاصی از سال دارای شیوع بیشتری می‌باشد و متغیر زمان به عنوان یکی از مهمترین مولفه در تشدید بیماری لپتوسپیروز بشمار می‌آید، قصد داریم تا توزیع بیماری مورد نظر را به صورت ماهیانه و زمانمند بررسی نماییم. در نهایت، پیشنهاد می‌شود که از دیگر روش‌های طبقه‌بندی و پیش‌بینی کننده مانند ماشین بردار پشتیبان (SVM) برای تولید نقشه پیش‌بینی توزیع مکانی- زمانی بیماری لپتوسپیروز، استفاده شود.

سپاسگزاری

نویسندگان این مقاله از تمامی پرسنل مرکز کنترل و پیشگیری بیماری‌های واگیر وزرات بهداشت، درمان و آموزش پزشکی ایران و به خصوص جناب آقای دکتر محمدرضا شیرزادی، رئیس مرکز کنترل و پیشگیری بیماری‌های واگیر بابت کمک‌ها و راهنمایی‌های علمی و پژوهشی‌شان و فراهم نمودن داده‌های بیماری لپتوسپیروز مورد استفاده در این تحقیق، کمال تشکر و قدردانی را دارند.

مراجع

- [1] M. C. Vega-Corredor and J. Opadeyi, (2014), "Hydrology and public health: linking human leptospirosis and local hydrological dynamics in Trinidad, West Indies," *Earth Perspectives*, vol. 1, pp. 1-14.
- [2] A. Rafiei, O. Amjadi, and F. Babamahmoodi, (2014), "Leptospirosis or Rice Field Fever: A review on the pathogenesis," *Journal of Clinical Excellence*, vol. 2, pp. 23-39.
- [3] H. Honarmand, S. Eshraghi, M. Khorramizadeh, R. Hartskeerl, F. Ghanaei, G. Abdolpour, et al., (2007), "Distribution of Human Leptospirosis in Guilan Province, Northern Iran," *Iranian J Publ Health*, vol. 36, pp. 68-72.
- [4] A. Rafiei, A. H. Zadeh-Omran, F. Babamahmoodi, R. A. Navaei, R. Valadan, and I. Sari, (2012), "Review of Leptospirosis in Iran," *Journal of Mazandaran University of Medical Sciences (JMUMS)*, vol. 22.
- [5] W. H. Organization, (2003), "Human leptospirosis: guidance for diagnosis, surveillance and control,".
- [6] P. Akhavan, M. Karimi, and P. Pahlavani, (2014), "Risk Mapping of Cutaneous Leishmaniasis via a Fuzzy C Means-based Neuro-Fuzzy Inference System," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, p. 19.
- [7] Bayatani A and Sadeghi A, (2012), "Spatial Analysis of Environmental Factors of Cutaneous Leishmaniasis in Iran Using GIS," *Hakim Research Journal*, vol. 15, pp. 158-165, (in Persian).
- [8] M. Azarmehr, M. Mesgari, and a. M. Karimi, (2010), "patio-temporal modeling of malaria using Geographic Information System and Cellular Automata method," *Iranian Journal of Infectious Diseases*, vol. 15, pp. 61-69, (in Persian).
- [9] M. Mesgari and Z. Masoomi, (2008), "GIS applications in public health as a decision making support system and it's limitation in Iran," *World appl Sci J (Supple 1)*, vol. 3, pp. 73-77.
- [10] P. Fine, C. G. Victora, K. J. Rothman, P. S. Moore, Y. Chang, V. Curtis, et al., (2013), "John Snow's legacy: epidemiology without borders," *The Lancet*, vol. 381, pp. 1302-1311.
- [11] S. Riedel, (2005), "Edward Jenner and the history of smallpox and vaccination," in *Baylor University Medical Center. Proceedings*, p. 21.

- [12] M. Castillo-Riquelme, Z. Chalabi, J. Lord, F. Guhl, D. Campbell-Lendrum, C. Davies, et al., (2008), "Modelling geographic variation in the cost-effectiveness of control policies for infectious vector diseases: The example of Chagas disease," *Journal of health economics*, vol. 27, pp. 405-426.
- [13] L. Rinaldi, V. Musella, A. Biggeri, and G. Cringoli, (2006), "New insights into the application of geographical information systems and remote sensing in veterinary parasitology," *Geospatial Health*, vol. 1, pp. 33-47.
- [14] Z. Pezeshki, M. Tafazzoli-Shadpour, A. Mansourian, B. Eshrati, E. Omidi, and I. Nejadqoli, (2012), "Model of cholera dissemination using geographic information systems and fuzzy clustering means: Case study, Chabahar, Iran," *Public health*, vol. 126, pp. 881-887.
- [15] M. Rajabi, A. Mansourian, and A. Bazmani, (2012), "Susceptibility mapping of visceral leishmaniasis based on fuzzy modelling and group decision-making methods," *Geospatial health*, vol. 7, pp. 37-50.
- [16] R. J. King, D. H. Campbell-Lendrum, and C. R. Davies, (2004), "Predicting geographic variation in cutaneous leishmaniasis, Colombia," *Emerging infectious diseases*, vol. 10, pp. 598-607.
- [17] M. Sedehi, Y. Mehrabi, A. Kazemnejad, and F. Hadaegh, (2010), "Comparison of artificial neural network, logistic regression and discriminant analysis methods in prediction of metabolic syndrome," *Iranian journal of endocrinology and metabolism*, vol. 11, pp. 638-646, 731, (in Persian).
- [18] C.-H. Lin and T.-H. Wen, (2011), "Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue," *International journal of environmental research and public health*, vol. 8, pp. 2798-2815.
- [19] M. Ali-Akbarpour, A. Mohammadbeigi, S. H. R. Tabatabaee, and G. Hatam, (2012), "Spatial analysis of eco-environmental risk factors of cutaneous leishmaniasis in southern Iran," *Journal of cutaneous and aesthetic surgery*, vol. 5, p. 30.
- [20] H. G. Zadeh, O. Pakdelazar, J. Haddadnia, G. Rezai-Rad, and M. Mohammad-Zadeh, (2012), "Diagnosing breast cancer with the aid of fuzzy logic based on data mining of a genetic algorithm in infrared images," *Middle East Journal of Cancer*, vol. 3, pp. 119-129.
- [21] R. Rismala, T. H. Liong, and A. Ardiyanti, (2013), "Prediction of malaria incidence in Banggai regency using evolving neural network," in *Technology, Informatics, Management, Engineering, and Environment (TIME-E)*, 2013 International Conference on, pp. 89-94.
- [22] M. Rajabi, A. Mansourian, P. Pilesjö, and A. Bazmani, (2014), "Environmental modelling of visceral leishmaniasis by susceptibility-mapping using neural networks: a case study in north-western Iran," *Geospatial health*, vol. 9, pp. 179-191.
- [23] P. Akhavan, M. Karimi, and A. Mansourian, (2014), "Prediction of Cutaneous Leishmaniasis Based on Artificial Neural Network" *Proc. of the 32nd National and the 1st International Geosciences Congress*, 16-19 February, (in Persian).
- [24] Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang, (2015), "Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China)," *Applied Soft Computing*, vol. 35, pp. 280-290.
- [25] R. Raghavan, K. Brenner, J. Higgins, J. S. Hutchinson, and K. Harkin, (2012), "Evaluations of hydrologic risk factors for canine leptospirosis: 94 cases (2002–2009)," *Preventive veterinary medicine*, vol. 107, pp. 105-109.
- [26] R. Raghavan, K. Brenner, J. Higgins, D. Van der Merwe, and K. R. Harkin, (2011), "Evaluations of land cover risk factors for canine leptospirosis: 94 cases (2002–2009)," *Preventive veterinary medicine*, vol. 101, pp. 241-249.
- [27] S. Zaki, W. Shieh, U. D. of Agriculture, C. f. D. Control, and Prevention, (1996), "Leptospirosis associated with outbreak of acute febrile illness and pulmonary haemorrhage, Nicaragua, 1995," *The Lancet*, vol. 347, pp. 535-536.
- [28] A. M. Nia, A. Alimohammadi, R. Habibi, and M. Shirzadi, (2015), "Spatial and Statistical Analysis of Leptospirosis in Guilan Province, Iran," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, p. 497.
- [29] I. T. Jolliffe and A. Philipp, (2010), "Some recent developments in cluster analysis," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, pp. 309-315.
- [30] A. Entezari, A. DadashiRudbari and M. Asadi, (2016), "Assessing the autocorrelation of spatial-temporal temperature change in heat islands of Khorasan Razavi Province," *The Journal of Geography and Environment Hazardz*, vol. 4, pp. 125-146, (in Persian).
- [31] B. D. Ripley, (2005), "Spatial statistics: John Wiley & Sons," vol. 575.
- [32] W. S. Kendall, (1998), "Perfect simulation for the area-interaction point process," in *Probability towards 2000*, ed: Springer, pp. 218-234.
- [33] A. Narayanan, (1993), "Probability and Statistics in Engineering and Management Science," *Technometrics*, vol. 35, pp. 238-239.

- [34] A. Mitchell, (1999), "The ESRI Guide to GIS Analysis: Geographic patterns & relationships," vol. 1: ESRI, Inc.
- [35] M. O. Ruiz, C. Tedesco, T. J. McTighe, C. Austin, and U. Kitron, (2004), "Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002," *International Journal of Health Geographics*, vol. 3, p. 1.
- [36] A. Getis, L. Anselin, A. Lea, M. Ferguson, and H. Miller, (2004), "Spatial analysis and modeling in a GIS environment," *A research agenda for geographic information science*. CRC, Boca Raton, FL, pp. 157-196.
- [37] A. Mitchell, (2005), "The ESRI guide to GIS analysis, Volume 2: Spatial Measurements and Statistics. Redlands," ed: CA: Esri Press.
- [38] A. Getis and J. K. Ord, (1992), "The analysis of spatial association by use of distance statistics," *Geographical analysis*, vol. 24, pp. 189-206.
- [39] J. Karami, S. Delfan, and A. Shamsoddini, (2016), "Role of Time in Spatial Analysis of Diseases in Tehran," *Journal of Geomatics Science and Technology*, vol. 5, pp. 227-238, (in Persian).
- [40] K. Ghorbani, (2012), "Geographically weighted regression: A method for mapping isohyets in Gilan Province," *Journal of Water and Soil*, Vol. 26, No. 3, pp. 743-752, (in Persian).
- [41] M. Karimi, Z. Akbari, and M. SaadiMesgari, (2014), "Development a Spatial Recommender System for Real Estate Marketing with Temporal Valuingcapability," *Journal of Geomatics Science and Technology*, vol. 3, pp. 41-54, (in Persian).
- [42] A. S. Fotheringham, C. Brunsdon, and M. Charlton, (2003), "Geographically weighted regression: the analysis of spatially varying relationships: John Wiley & Sons,".
- [43] M. Halimi, M. Delavari, and a. M. JaafariMadrak, (2013), "Spatial Modeling of Annual Parasite Incidence of Malaria in Sistan and Balouchistan by Province Geographical Weighted Regression," *The Journal of Spatial Planning*, vol. 17, pp. 85-108, (in Persian).
- [44] R. Taghizadeh, F. Sarmadian, G. Savaghebi, M. Omid, N. Toomanian, M. Rousta, et al., (2013), "Comparison of Neuro-Fuzzy, Genetic Algorithm, Artificial Neural Network and Multivariate Regression for Prediction of Soil Salinity (Case study: Ardakan City)," *The Journal of Range and Watershed Management*, vol. 66, pp. 207-222, (in Persian).
- [45] S. Mohammady and M. R. Delavar, (2014), "Urban Expansion Modeling with Logistic Regression," *Journal of Geomatics Science and Technology*, vol. 4, pp. 77-86.
- [46] D. M. Skapura, (1996), "Building neural networks: Addison-Wesley Professional,".
- [47] L. Zhang and Y. Zhen, (2006), "An artificial neural network model of the landscape pattern in Shanghai metropolitan region, China," *Frontiers of Biology in China*, vol. 1, pp. 463-469.
- [48] A. Tayyebi, B. C. Pijanowski, and A. H. Tayyebi, (2011), "An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran," *Landscape and Urban Planning*, vol. 100, pp. 35-44.
- [49] N. Ritter, T. Logan, and N. Bryant, (1988), "Integration of neural network technologies with geographic information systems," in *Proceedings of the GIS symposium: integrating technology and geoscience applications*, pp. 102-103.
- [50] D. Panagoulia, (2006), "Artificial neural networks and high and low flows in various climate regimes," *Hydrological sciences journal*, vol. 51, pp. 563-587.
- [51] X. Hu and Q. Weng, (2009), "Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks," *Remote Sensing of Environment*, vol. 113, pp. 2089-2102.
- [52] R. J. Hyndman and A. B. Koehler, (2006), "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, pp. 679-688.
- [53] C. Tofallis, (2015), "A better measure of relative prediction accuracy for model selection and model estimation," *Journal of the Operational Research Society*, vol. 66, pp. 1352-1362.
- [54] A. Mollalo, A. Alimohammadi, M. Shirzadi, and M. Malek, (2015), "Geographic Information System-Based Analysis of the Spatial and Spatio-Temporal Distribution of Zoonotic Cutaneous Leishmaniasis in Golestan Province, North-East of Iran," *Zoonoses and public health*, vol. 62, pp. 18-28.
- [55] M. H. Kutner, C. Nachtsheim, and J. Neter, (2004), "Applied linear regression models: McGraw-Hill/Irwin,".
- [56] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, (2003), "A practical guide to support vector classification,".
- [57] N. Draper and H. Smith, (1998), "Applied regression analysis: Wiley Interscience," New York, pp. 505-553.
- [58] B. C. Pijanowski, S. Pithadia, B. A. Shellito, and K. Alexandridis, (2005), "Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States," *International Journal of Geographical Information Science*, vol. 19, pp. 197-215.
- [59] S. Sousa, S. Caeiro, and M. Painho, (2002), "Assessment of map similarity of categorical maps using Kappa statistics," *ISEGI, Lisbon*.