

مدل‌سازی و تحلیل روند توزیع بیماری لیشمانیوز با استفاده از شبکه عصبی پرسپترون چندلایه و رگرسیون بردار پشتیبان (مطالعه موردی: روستاهای استان اصفهان)

نگار شعبانپور^{1*}، ندا کفاش چرندابی²، محمدرضا شیرزادی³

¹ کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین

طوسی

n_shabanpour@yahoo.com

² استادیار گروه نقشه‌برداری - دانشکده فنی مهندسی مرن - دانشگاه تبریز

n_kaffash@tabrizu.ac.ir

³ رئیس مرکز کنترل و پیشگیری بیماری‌های واگیردار وزارت بهداشت - درمان و آموزش پزشکی

shirzadim@gmail.com

(تاریخ دریافت: اردیبهشت 1400، تاریخ تصویب: مرداد 1401)

چکیده

روستاهای واقع در استان اصفهان یکی از مناطق مستعد شیوع بیماری لیشمانیوز است که با وقوع زخم روی پوست مشخص می‌شود. برای پیش‌بینی شیوع بیماری لیشمانیوز در آینده، نظارت مداوم بر روند توزیع مکانی این بیماری ضروری است. مدلسازی بیماری لیشمانیوز با استفاده از دو الگوریتم یادگیری ماشین به نام‌های رگرسیون بردار پشتیبان (SVR) و شبکه عصبی پرسپترون چندلایه (MLP) انجام شده‌است. تحلیل نتایج نشان می‌دهد الگوریتم SVR با $RMSE=0/170$ و $NRMSE=12/08$ در مقایسه با MLP با مقدار $RMSE=0/348$ و $NRMSE=11/45$ عملکرد بهتری دارد. پارامترهای مورد استفاده در تحقیق شامل دمای هوا، رطوبت، بارش، ارتفاع، سرعت باد، جمعیت و نرخ تراکم بیماری لیشمانیوز فرایند مدلسازی مورد استفاده قرار گرفته‌است، که از این تعداد (70%) برای آموزش مدل و (30%) باقی‌مانده برای ارزیابی مدل مورد استفاده قرار گرفته‌است. نتایج شاخص تحلیل مکانی نشان داد که الگوی توزیع مکانی بیماری لیشمانیوز در سال‌های 1397 تا 1399 بصورت خوشه‌ای بوده‌است و پارامترهای رطوبت و بارش بیشترین اثر را بر روی شیوع بیماری لیشمانیوز در این منطقه داشته‌است.

واژگان کلیدی: بیماری لیشمانیوز، شبکه عصبی پرسپترون چندلایه، رگرسیون بردار پشتیبان

* نویسنده رابط

1- مقدمه

لیشمانیوز یک مشکل بهداشتی عمده و رو به رشد در بسیاری از کشورهای جهان است که در آن عوامل خطر محیطی و انسانی در حال افزایش هستند [1]. لیشمانیوز نام گروهی از بیماری‌های انگلی می‌باشد که به اشکال گوناگونی از جمله احشایی (VL)، جلدی سالک (CL) و مخاطی-جلدی (ML) ظاهر می‌شود [2]. سالک شایع‌ترین فرم بیماری لیشمانیوز است که از طریق نیش پشه خاکی فلботوموس ماده منتقل می‌شود [3] و به عنوان یک بیماری مشترک بین انسان و حیوان، مشکل مهم بهداشت عمومی و سلامت در سراسر جهان است [4]. این بیماری با ضایعات پوستی مشخص می‌شود که بطور معمول در چند هفته یا چند ماه پس از قرار گرفتن در معرض نور ظاهر و منجر به زخم شدید می‌شود [3]. که زمان زیادی طول می‌کشد التیام یابد و به دنبال آن مشکلات جدی اجتماعی-اقتصادی را به جامعه وارد می‌کند [5]. با توجه به ماهیت پویای بیماری سالک به عنوان یک بیماری همه گیر از طریق ناقل، عوامل اقلیمی و محیطی و نیز عوامل جمعیتی و فعالیت انسانی به طور قابل توجهی تحت تاثیر این بیماری قرار گرفته‌اند [4].

سالانه بیش از 24000 مورد این نوع سالک در کشور ایران رخ می‌دهد [4] و به متوسط در 17 استان این کشور همه گیر و بسیار شایع می‌باشد [6]. استان اصفهان که در مرکز ایران قرار دارد، از دیرباز به عنوان یکی از مهم‌ترین مناطق همه‌گیر بیماری سالک شناخته شده‌است و هنوز هم تعداد زیادی از عفونت‌های سالانه را با حدود 2200 مورد در سال 1392 تجربه می‌کند. این بیماری ناشی از لیشمانیا ماژور و تک‌تاخته‌ی لیشمانیا است، که در حال حاضر در اصفهان شیوع دارد [1] و همچنین مطالعه تحقیق [7] نشان داده‌است که روند شیوع بیماری سالک در سال‌های اخیر بصورت افزایشی بوده‌است و نیاز است راه کارهایی جهت کاهش و کنترل این سالک انجام شود. بر این اساس، در این تحقیق برای مدلسازی بیماری سالک، محل اقامت بیماران، روستاهای واقع در استان اصفهان مورد ارزیابی قرار گرفت.

سامانه اطلاعات مکانی با توجه به کاربردهای مفیدی که در ثبت و ذخیره‌سازی، تجزیه و تحلیل، مدیریت و نمایش اطلاعات بروز بیماری‌ها را دارد، می‌تواند برای الگوی توزیع

مکانی سالک، بررسی عوامل احتمالی موثر بر شیوع سالک، بررسی روابط مکانی-زمانی و پیش‌بینی مکانی بیماری سالک مورد استفاده قرار گیرد، زیرا زمانی که یک بیماری تحت تاثیر پارامترهای محیطی قرار می‌گیرد، داده‌های سنجش‌ازدور و GIS تا حد زیادی به مدیریت و یافتن الگوی شیوع آن بیماری کمک می‌کند [8] و در کنار آن ابزارهایی را برای ارزیابی استراتژی‌های مختلف، به منظور جلوگیری از شیوع آن فراهم کند [9]. همچنین GIS می‌تواند ویژگی‌های مختلف محیطی، اکولوژیکی و اطلاعات مکانی را یکپارچه کند تا اثرات آن‌ها را بر روی شیوع بیماری ارزیابی کند [4]. بنابراین، ادغام GIS، داده‌های سالک و عوامل محیطی موثر بر آن می‌تواند به پیش‌بینی گسترش، مدیریت و کنترل شیوع بیماری کمک کند [10].

با توجه به این‌که، تعداد بیمارانی که برای درمان‌های مختلف به بیمارستان‌ها مراجعه می‌کنند در حال افزایش است، همواره حجم زیادی از اطلاعات بیماران در پایگاه داده‌های پزشکی جمع‌آوری می‌شود، اما مشکل عدم استفاده از این منابع است. در همین راستا روش‌های یادگیری ماشین می‌تواند به پزشکان در زمینه پیش‌بینی و تشخیص بیماری‌ها نقش موثری ایفا کند. از طرفی، عوامل و پارامترهای نامشخصی در شیوع بیماری سالک موثر است، که رابطه بین این پارامترها و سالک ممکن است پیچیده و حتی غیرخطی باشد [11]. به همین علت، محبوبیت رو به رشد الگوریتم‌های یادگیری ماشین را می‌توان به توانایی آن‌ها در تقریب‌زدن هر رابطه عملکردی غیرخطی پیچیده نسبت داد [12]. در این تحقیق توزیع مکانی-زمانی سالک، تعیین مناطق پرخطر سالک و ارتباط آن با عوامل محیطی با استفاده از سیستم اطلاعات جغرافیایی بررسی گردید و در نهایت مدلسازی بیماری سالک با استفاده از دو الگوریتم یادگیری ماشین شامل روش شبکه عصبی پرسپترون چند لایه و رگرسیون بردار پشتیبان انجام شد.

2- پیشینه تحقیق

بیماری‌ها ممکن است مرتبط با آلودگی محیطی، مرتبط با رفتارهای فردی یا گروهی و یا مرتبط با زمینه ژنتیکی باشند [13]. شیوع بیماری سالک به صورت مستقیم و یا غیر مستقیم تحت تاثیر عوامل محیطی

می‌باشد [14]. در جهان تحقیقات زیادی در ارتباط به بررسی بیماری سالک انجام گرفته‌است.

تحقیق [15] در ایتالیا انجام گرفت. در این مطالعه حضور پشه‌های خاکی ناقل بیماری سالک در سه سال متوالی بررسی شد. یکی از نتایج این تحقیق مربوط به تاثیر شرایط آب و هوایی بر شیوع پشه‌های خاکی است، اما در این مطالعه تنها پارامترهای رطوبت و دما بررسی شده‌است [16]. نیز در استان گلستان انجام شده‌است. در این مطالعه از محیط شبیه‌سازی NetLogo استفاده گردید. علاوه بر این، شرایط آب و هوایی منطقه مطالعه بررسی نشده و تنها زیستگاه‌های ناقلین بیماری سالک شامل مزارع کشاورزی و اطراف رودخانه‌ها مطالعه شده‌است. تحقیق [2] در استان گلستان انجام گرفت. در این تحقیق پارامترهای محیطی تاثیرگذار بر بیماری شامل دما، بارش، شیب، ارتفاع، پوشش گیاهی، نزدیکی به محل نگهداری دام و محدوده‌های حضور ناقلین بیماری سالک می‌باشند. در این مطالعه به منظور مدل‌سازی بیماری سالک تنها از مدل رگرسیون لجستیک استفاده شده‌است. در تحقیق [17] که در شهرستان‌های استان بوشهر انجام گرفت. پارامترهای تعداد بیماران، اطلاعات مربوط به سن و جنسیت بیمار، تعداد زخم‌های ناشی از گزش ناقلین بیماری و موقعیت آن‌ها بر روی بدن فرد بررسی شد. در این مطالعه مدل‌سازی و تاثیر پارامترهای محیطی بر روند شیوع بیماری بررسی نشده‌است. تحقیق [18] در استان گلستان انجام گرفت. در این مطالعه از الگوریتم‌های رگرسیون و زنده‌دار جغرافیایی و شبکه عصبی پرسپترون چند لایه استفاده شد. پارامترهای محیطی تاثیرگذار بر بیماری سالک شامل میانگین دمای هوا، میانگین رطوبت، مجموع بارش سالیانه، ارتفاع، شیب، جهت شیب و تعداد روزهای یخبندان می‌باشند. در این مطالعه پارامتر جمعیت و روند توزیع مکانی بیماری نیز تعیین نشده‌است. در مطالعه [19] تنها توزیع مکانی-زمانی بیماری سالک بررسی گردید و نشان دادند که شیوع سالک در ارتفاعات پایین و با آب و هوای خشک و نیمه‌خشک و در شرایط اقتصادی-اجتماعی ضعیف واقع شده‌است. در این تحقیق مدل‌سازی بیماری سالک وجود نداشته و تعداد محدودی از پارامترهای محیطی در نظر گرفته شده‌است. [20] توزیع مکانی بیماری سالک را در استان بوشهر بررسی کرد. در این مطالعه سن و جنسیت بیماران در نظر گرفته شد. نتایج تحقیق نشان داد بیماری سالک در مردان در

سنین 10 تا 30 سال وجود دارد. در این مطالعه تاثیر پارامترهای محیطی و مدل‌سازی شیوع بیماری سالک بررسی نشده‌است. تحقیق دیگری در استان گلستان توسط [21] انجام گرفت. در این مطالعه پارامترهای محیطی موثر بر بیماری سالک دمای هوا، بارش، شیب و فاصله تا نزدیک‌ترین رودخانه در نظر گرفته شد. در این تحقیق پارامترهای رطوبت و جمعیت در این مطالعه بررسی نشده‌است، زیرا که این پارامترها تاثیر زیادی در شیوع بیماری سالک دارند. همچنین، در این مطالعه تعداد بیماران بررسی نشد و هدف مطالعه تنها بررسی پراکندگی پشه‌های خاکی ناقل بیماری سالک بوده‌است. در مطالعه [11] در استان گلستان از پارامترهای محیطی موثر بر بیماری سالک شامل جمعیت، رطوبت نسبی، بارش، دمای هوا، پوشش گیاهی و ارتفاع استفاده شد. پارامتر تراکم بیماری سالک بر اساس جمعیت و تعداد بیماران منطقه در روند مدل‌سازی بیماری سالک بسیار تاثیرگذار است که در این تحقیق در نظر گرفته نشده‌است و به منظور مدل‌سازی بیماری تنها از الگوریتم شبکه عصبی مصنوعی استفاده است. این تحقیق برای مدل‌سازی بیماری سالک از الگوریتم‌های بردار پشتیبان و شبکه عصبی مصنوعی استفاده کرد؛ و این مدل‌سازی با در نظر داشتن پارامترهای محیطی موثر بر بیماری سالک در منطقه مطالعاتی شامل دمای هوا، رطوبت، بارش، ارتفاع، سرعت باد، جمعیت و نرخ تراکم بیماری سالک در سطح روستاهای استان اصفهان انجام شد.

3- منطقه مورد مطالعه

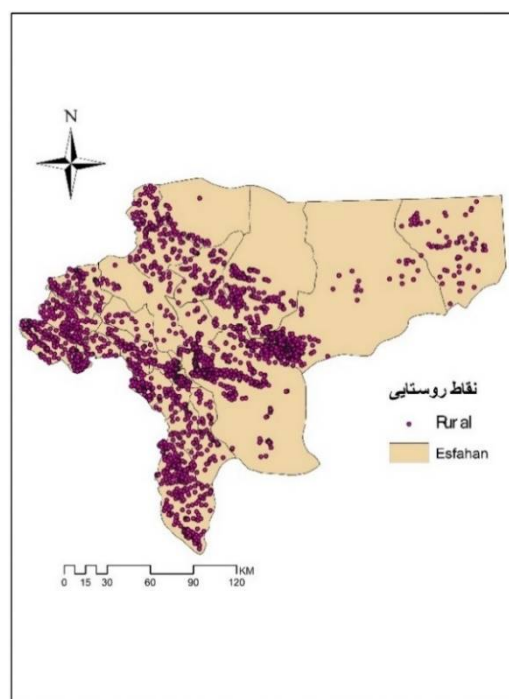
استان اصفهان در مرکز ایران با مساحتی حدود 106786 کیلومتر مربع در عرض جغرافیایی 30 درجه و 43 دقیقه و 27 درجه و 34 دقیقه شمالی و طول جغرافیایی 49 درجه و 36 دقیقه و 55 درجه و 31 دقیقه شرقی قرار دارد. ارتفاع 1500 متر بالاتر از سطح دریا، متوسط درجه حرارت سالانه 16/7 درجه سانتی‌گراد و متوسط میزان بارندگی سالانه آن 116/9 میلی‌متر است. روستاهای مورد مطالعه در استان اصفهان در شکل (1) نشان داده شده‌است.

4- جمع‌آوری و آماده‌سازی داده‌ها

در این تحقیق اطلاعات مربوط به تعداد بیماران مبتلا به بیماری سالک در هر روستا در سطح استان اصفهان در

بازه زمانی 1397 تا 1399 از مرکز مدیریت بیماری‌های واگیردار جمع‌آوری گردید. سپس از طریق داده‌های جمع‌آوری شده بدلیل وابستگی آن به جمعیت، نرخ بیماری سالک بر اساس معادله (1) بدست آمد [12]. در مرحله بعد، به منظور انجام تجزیه و تحلیل، اطلاعات وارد پایگاه داده مکانی در نرم‌افزار ArcGIS 10.3 شد.

$$(1) \quad \text{نرخ بیماری سالک} = \frac{\text{تعداد بیماری سالک}}{\text{جمعیت}}$$



شکل 1- موقعیت روستاهای مورد مطالعه در استان اصفهان [سالنامه آماری، 1390]

پارامترهای محیطی در دسترس شامل سرعت باد، بارش، رطوبت، دمای هوا و ارتفاع می باشد. این پارامترها از 21 ایستگاه سینوپتیک در استان اصفهان به صورت سالیانه [20]، همچنین اطلاعات مربوط به مقادیر ارتفاعی از مدل رقومی ارتفاع (DEM) با اندازه شبکه 30 متر استخراج گردید. برای آماده‌سازی پارامترهای محیطی از تکنیک درونیابی IDW با اندازه شبکه 30 متر در نرم‌افزار ArcGIS 10.3 استفاده گردید.

5- روش انجام تحقیق

در این تحقیق هدف ما بررسی روند شیوع و شناسایی الگوی مکانی بیماری سالک در سطح روستاهای استان

اصفهان است. در ابتدا پارامترهای محیطی از سال 1397 تا سال 1399 جمع‌آوری شدند و در نرم‌افزار ArcGIS آماده‌سازی شدند. داده‌های مربوط به نرخ بیماری سالک نیز در نرم‌افزار ArcGIS وارد گردید و با استفاده از تابع تخمین تراکم کرنل به نقشه رستری تراکم سالک به عنوان پارامتر وابسته در مدل‌سازی آماده‌سازی گردید. سپس، با استفاده از آزمون آماری پیرسون، موثرترین داده‌های محیطی انتخاب شدند و به عنوان پارامترهای مستقل وارد فرایند مدل‌سازی شدند. در این مطالعه، از الگوریتم‌های یادگیری ماشین شامل رگرسیون بردار پشتیبان و شبکه عصبی پرسپترون چند لایه برای مدل‌سازی بیماری سالک استفاده گردید و در نهایت دقت هر یک از الگوریتم‌ها در مرحله ارزیابی و آموزش مدل بررسی شد. شکل (2) مراحل انجام کار را بصورت فلوچارت نشان می‌دهد.

5-1- خودهمبستگی مکانی

خود همبستگی مکانی به همبستگی بین متغیرهای مشابه در موقعیت‌های مکانی مختلف و روابط بین یک مقدار متغیر مکانی و مقادیر نزدیک آن اشاره دارد. خودهمبستگی مکانی از شاخص‌های جهانی و محلی استفاده می‌کند. این تحلیل آماری جهانی توزیع کلی یک متغیر را به منظور تعیین اینکه آیا خوشه‌های مکانی این متغیر در یک ناحیه بزرگتر وجود دارد یا خیر توصیف می‌کند و از یک مقدار واحد برای انعکاس درجه خودهمبستگی ناحیه استفاده می‌کند، در مقابل، خودهمبستگی محلی واحدهای مکانی که در همسایگی با یکدیگر هستند را بررسی می‌کند [21]. در واقع، شاخص‌های جهانی توزیع رویدادها را ارزیابی می‌کند و مکان خوشه‌ها را مشخص نمی‌کنند [9]. بنابراین، در این تحقیق، از شاخص‌های موران جهانی، $Getis - ord Gi^*$ ، برای بررسی خود همبستگی مکانی و همچنین شناسایی الگوی مکانی بیماری سالک استفاده گردید.

5-1-1- شاخص موران جهانی

وجود خودهمبستگی مکانی برای هر سال مطالعه در این تحقیق با استفاده از شاخص موران مورد آزمون قرار گرفت. موران یک شاخص رایج برای توصیف خودهمبستگی مکانی است. مقدار این شاخص بین $[-1, +1]$ متغیر است. مقادیر $+1$ نشان‌دهنده خودهمبستگی مکانی مثبت و مقادیر

$$I = \frac{\sum_{j=1}^n W_{ij} X_j - \bar{X} \sum_{j=1}^n W_{ij}}{S \sqrt{\frac{n \sum_{j=1}^n W_{ij}^2 - (\sum_{j=1}^n W_{ij})^2}{n-1}}} \quad (3)$$

در این فرمول X_j مقدار ویژگی برای نقطه j است، W_{ij} وزن مکانی بین نقطه j و i است، n برابر با تعداد کل نقاط است.

5-2- تحلیل آماری

تحلیل‌های آماری مجموعه‌ای از روش‌های آماری می‌باشند که به بررسی ارتباط میان پارامترهای مستقل و پارامتر وابسته می‌پردازند. در این تحقیق به منظور بررسی ارتباط میان پارامترهای محیطی و سالک از آزمون آماری پیرسون استفاده شده است.

5-2-1- آزمون آماری پیرسون

ضریب همبستگی پیرسون یک شاخص متقارن است که مقدار قدر مطلق آن، شدت یا درجه رابطه بین دو پارامتر و علامت آن (مثبت یا منفی) جهت رابطه (مستقیم یا معکوس) را نشان می‌دهد. مقدار این ضریب همیشه بین $[-1, +1]$ می‌باشد. زمانی که ارتباط میان پارامترها تاثیر معکوس بر روی یکدیگر داشته باشند مقدار این همبستگی بین $[-1, 0]$ است و هنگامی که ارتباط میان پارامترها مستقیم باشد، افزایش یک پارامتر باعث افزایش پارامتر دیگر می‌شود و مقدار این پارامتر در بازه $[0, +1]$ قرار می‌گیرد. مقدار صفر بیان می‌کند که هیچگونه ارتباطی میان پارامترها وجود ندارد. ضریب همبستگی پیرسون بر اساس رابطه (4) محاسبه می‌گردد [25]

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \cdot \sum (Y - \bar{Y})^2}} \quad (4)$$

محاسبه ضریب همبستگی پیرسون از طریق همبستگی بین پارامترهای X ، Y و انحراف از میانگین $\bar{X}$ و $\bar{Y}$ می‌باشد.

5-3- تخمین تراکم کرنل بیماری سالک

تابع (KDE^1) یک روش غیرپارامتری است که می‌تواند برای بررسی چگالی مکانی داده‌های نقطه‌ای مورد استفاده

قرار گیرد [26]. در این روش هر چه مقادیر مشاهدات به نقطه x نزدیک‌تر باشند، وزن بیشتری برای آن نقاط در نظر گرفته می‌شود. بنابراین، مقدار این تابع در نقطه x با استفاده از رابطه (5) محاسبه می‌شود:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right) \quad (5)$$

مقدار پارامتر h مربوط به شعاع جستجو، k مربوط به تابع کرنل و X_i مربوط به مقادیر مشاهدات در شعاع جستجو می‌باشد. مقدار برآورد کرنل در نقطه x ، میانگین مقدار n کرنل در آن نقطه است. به عبارت دیگر، تابع کرنل به داده‌ها در همسایگی یک نقطه خاص وزن $\frac{1}{n}$ را اختصاص می‌دهد. مقدار h مقدار هموارسازی تابع نیز نامیده می‌شود که مقدار آن میزان هموار بودن در برآورد کرنل را نشان می‌دهد. بنابراین، هر چه مقدار این پارامتر کوچک‌تر در نظر گرفته شود، منحنی حاصل از برآورد تابع کرنل ناهموارتر شده و جزئیات مربوط به تراکم سالک بهتر نمایش داده می‌شود. در مقابل، هر چه میزان شعاع جستجو بزرگ‌تر در نظر گرفته شود، منحنی هموار تر و از نمایش جزئیات تابع تراکم کاسته می‌گردد [27]. در این تحقیق جهت برآورد تراکم بیماری سالک از نرم‌افزار ArcGIS استفاده گردید.

5-4- مدلسازی بیماری سالک

الگوریتم‌های یادگیری ماشین، مدل‌های ریاضی هستند که می‌توانند رابطه بین داده‌های ورودی و خروجی هدف را با دریافت داده‌های آموزشی ترسیم کنند [28]. این فرایند شامل دو مرحله است که به عنوان مرحله یادگیری یا آموزش و مرحله آزمایش یا پیش‌بینی شناخته می‌شود. هدف از آموزش، یادگیری رفتار تابع هدف برای برآورد مقادیر پیش‌بینی می‌باشد [17]. پس از آن یک مدل آموزش‌دیده می‌تواند متغیر مورد نظر را برای یک مجموعه داده مستقل جدید پیش‌بینی کند [1]. در این مطالعه از دو روش یادگیری ماشین شامل شبکه عصبی پرسپترون چند لایه و رگرسیون بردار پشتیبان استفاده گردید.

1 Kernel Density Estimation

5-4-1- شبکه عصبی پرسپترون چند لایه¹

شبکه عصبی یک مدل محاسباتی است که به منظور یافتن الگوی داده‌ها و تعمیم آن‌ها طراحی شده است [29]. شبکه عصبی تلاش می‌کند تا توانایی تصمیم‌گیری انسان را بصورت مصنوعی شبیه‌سازی کند. الگوریتم شبکه عصبی مصنوعی، واحد اصلی مغز انسان، نورون و اتصالات بین آن‌ها را شبیه‌سازی می‌کند و به آن‌ها اجازه می‌دهد تا با یکدیگر کار کنند و اطلاعات بدست‌آمده از تجربه عملکرد را ذخیره کند.

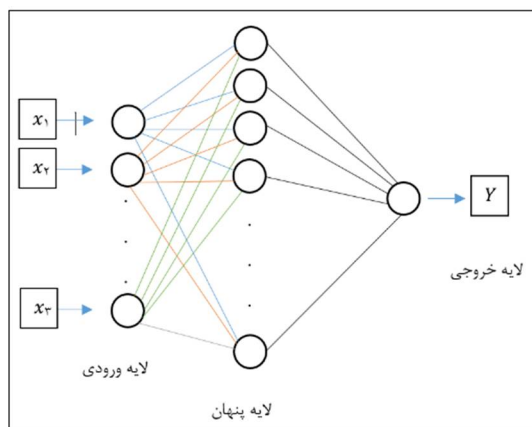
این سیستم‌ها به دانش قبلی از رابطه بین پارامترها نیاز دارند. مزیت الگوریتم شبکه عصبی مصنوعی توانایی آن برای یادگیری از موارد واقعی و یادگیری مجدد زمانی است که داده‌های جدید وارد سیستم می‌شوند [30].

ارتباط بین نورون‌ها می‌تواند دارای یک وزن عددی باشد که بر اساس تجربه تنظیم شده است و همچنین با ایجاد یک شبکه عصبی تطبیقی، امکان یادگیری برای شبکه فراهم می‌شود. در نتیجه می‌تواند وزن‌ها را تنظیم کند. یک شبکه عصبی معمولاً توسط سه پارامتر تعریف می‌شود: (1) الگوی اتصال بین لایه‌های مختلف نورون، (2) فرآیند یادگیری برای بروز رسانی وزن اتصالات، (3) تابع فعالسازی که مقادیر یک نورون مصنوعی را بر اساس مقادیر ورودی‌اش تعیین می‌کند. تعداد نورون‌های لایه‌های پنهان با هدف به حداقل رساندن میزان خطا بین خروجی‌های شبکه و خروجی‌های واقعی در مرحله آموزش و در مرحله آزمایش انتخاب می‌شوند [30]. در این مطالعه برای تنظیم بهینه این مقادیر از روش Grid Search استفاده شد.

$$f(x) = K \left(\sum_i w_i g_i(x) \right) \quad (6)$$

رابطه (6) به بیان ریاضی تابع شبکه عصبی $f(x)$ به صورت ترکیبی از توابع دیگر $g_i(x)$ تعریف می‌شود. معمولاً از مجموع وزن‌دار غیر خطی در شبکه عصبی استفاده می‌شود. که در آن K (معمولاً به تابع فعالسازی بر می‌گردد)، می‌تواند تابع تانژانت هیپربولیک، سیگموئید، خطی، تابع پله‌ای و گوسی باشد [31].

تابع فعالسازی یا همان تابع انتقال، یک تابع ریاضی است که بر اساس نوع آن بر ورودی‌های شبکه تاثیر گذاشته و سپس خروجی سلول عصبی را تولید می‌کند [30]. الگوریتم شبکه عصبی پرسپترون چندلایه، به عنوان یکی از انواع روش‌های شبکه عصبی مصنوعی مقید به فرض‌های اولیه توزیع داده‌ها نمی‌باشد و یکی از پر کاربرد ترین نوع شبکه‌های عصبی بشمار می‌آید، به طوری‌که نسبت به روش‌های دیگر بیشتر برای مدل‌سازی یک رویداد همچون مدل‌سازی پیش‌بینی بیماری کاربرد دارد. به منظور آموزش شبکه‌های عصبی پرسپترون چند لایه با هدف تنظیم پارامترهای شبکه و نزدیکی پاسخ شبکه به پاسخ واقعی، معمولاً از قانون یادگیری پس انتشار خطا استفاده می‌شود [2]. الگوریتم پس انتشار خطا برای کاهش گرادیان کاهش جهت حداقل کردن کل مربعات خطای خروجی است. هنگامی که خطایی در خروجی شبکه‌های پیشخور چند لایه‌ای رخ می‌دهد این خطا باید به سمت ورودی برگشت داده شود و باعث اصلاح وزن‌ها گردد. شکل (3) نمایی از شبکه عصبی پرسپترون چند لایه را نشان می‌دهد. همانطور که در شکل مشخص است، پارامترها بعنوان لایه ورودی، تعداد لایه پنهان و تعداد نورون‌ها با توجه به پیچیدگی مسئله متفاوت است.



شکل 3- شبکه پرسپترون چند لایه

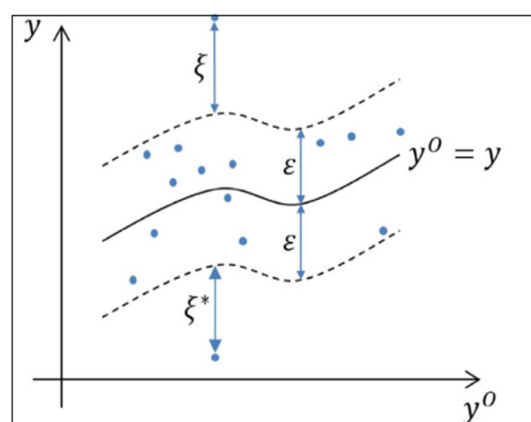
5-4-2- رگرسیون ماشین بردار پشتیبان

SVR^2 یک نوع سیستم یادگیری نظارت شده است که هم برای گروه‌بندی و هم برای تخمین و برآورد تابع برازش

2 Support Vector Regression (SVR)

1 Multi-layer Perceptron

داده‌ها در مسائل رگرسیون به کار می‌رود، به طوری که کمترین خطا در گروه‌بندی داده‌ها یا تابع برازش رخ دهد. هدف این مدل که مدل رگرسیونی SVM می‌باشد این است که تابع $f(x)$ را برای الگوهای آموزشی x طوری تشخیص دهد که بیشینه حاشیه را از مقادیر آموزشی y داشته باشد. به عبارت دیگر، مدل SVR مدلی است که منحنی با ضخامت ε را به داده‌ها برازش می‌دهد به نحوی که کمترین خطا در داده‌های آزمون صورت گیرد. بردارهای پشتیبان مجموعه‌ای از نقاط در فضای n بعدی داده‌ها هستند که مرز دسته‌ها را مشخص می‌کنند و مرزبندی و دسته‌بندی داده‌ها براساس آن‌ها انجام می‌شود و با جابه‌جایی یکی از آن‌ها خروجی دسته‌بندی ممکن است تغییر کند. یکی از کاربردهای ماشین‌بردار پشتیبان، رگرسیون ماشین‌بردار پشتیبان است که بطور کلی شامل دو مرحله آموزش و تست می‌باشد. کیفیت مدل‌های ماشین‌بردار پشتیبان وابسته به تنظیم مناسب پارامترهای مدل‌سازی می‌باشد [32]. بردار پشتیبان یک روش بدون مدل است که راه‌حل‌های موثری را برای حل مسائل بدون هیچ فرضی در مورد توزیع و وابستگی متقابل داده‌ها فراهم می‌کند [33].



شکل 4- نمایی از نحوه انجام بردار پشتیبان [34]

در حقیقت وپنیک¹ برای کاربرد ماشین‌بردار پشتیبان در مسائل رگرسیون از تابع خطای جدیدی استفاده کرد و خطاهایی را که در یک فاصله معین از مقادیر واقعی می‌باشند نادیده می‌گیرد. این موضوع بدان معناست که باید مقداری انحراف از ε را قابل قبول دانست، که در نهایت بر اساس اصل کمینه‌سازی خطای ساختاری، محدوده خطا بوسیله مسئله بهینه‌سازی بهینه می‌گردد

1 Vapnik

(شکل 4). مهم‌ترین مرحله در این الگوریتم، انتخاب صحیح تابع کرنل می‌باشد. تابع ریاضی مورد استفاده برای تبدیل SVR به عنوان تابع کرنل شناخته می‌شود، که در واقع با استفاده از این تابع داده‌هایی که بصورت خطی تفکیک پذیر هستند جداسازی می‌شوند. در واقعیت امر، اکثر داده‌ها بصورت خطی تفکیک‌پذیر نمی‌باشند، به همین ترتیب به منظور نگاشت داده‌ها به ابعاد بالاتر نیاز است از توابع کرنل استفاده شود [34]

6- ارزیابی نتایج

6-1- ارزیابی توزیع بیماری سالک

همانطور که از نتایج جدول (1) مشخص است، توزیع بیماری سالک در هر سه سال مطالعه بصورت خوشه‌ای است به این معنا که مقادیر مشابه در کنار یکدیگر واقع شده‌اند و خودهمبستگی مکانی مثبت می‌باشد. علاوه بر این، در سه سال مقدار شاخص P-value مقدار در نظر گرفته‌شده برای سطح معنی داری 0/05 کمتر می‌باشد، در نتیجه، فرض صفر مبنی بر تصادفی بودن توزیع سالک رد می‌شود. در شکل (5) با استفاده از شاخص Getis-ord G_i^* تحلیل نقاط داغ انجام شده‌است. برای هر مکان، امکان محاسبه شباهت آن با همسایگان و همچنین ارزیابی اهمیت آن فراهم شده‌است. مکان‌هایی با همسایگی با مقادیر شاخص بالا، که بعنوان مناطق داغ شناسایی می‌شود در قسمت‌های شمال غربی، مرکزی و جنوب شرقی استان در هر سه سال مشاهده می‌شود.

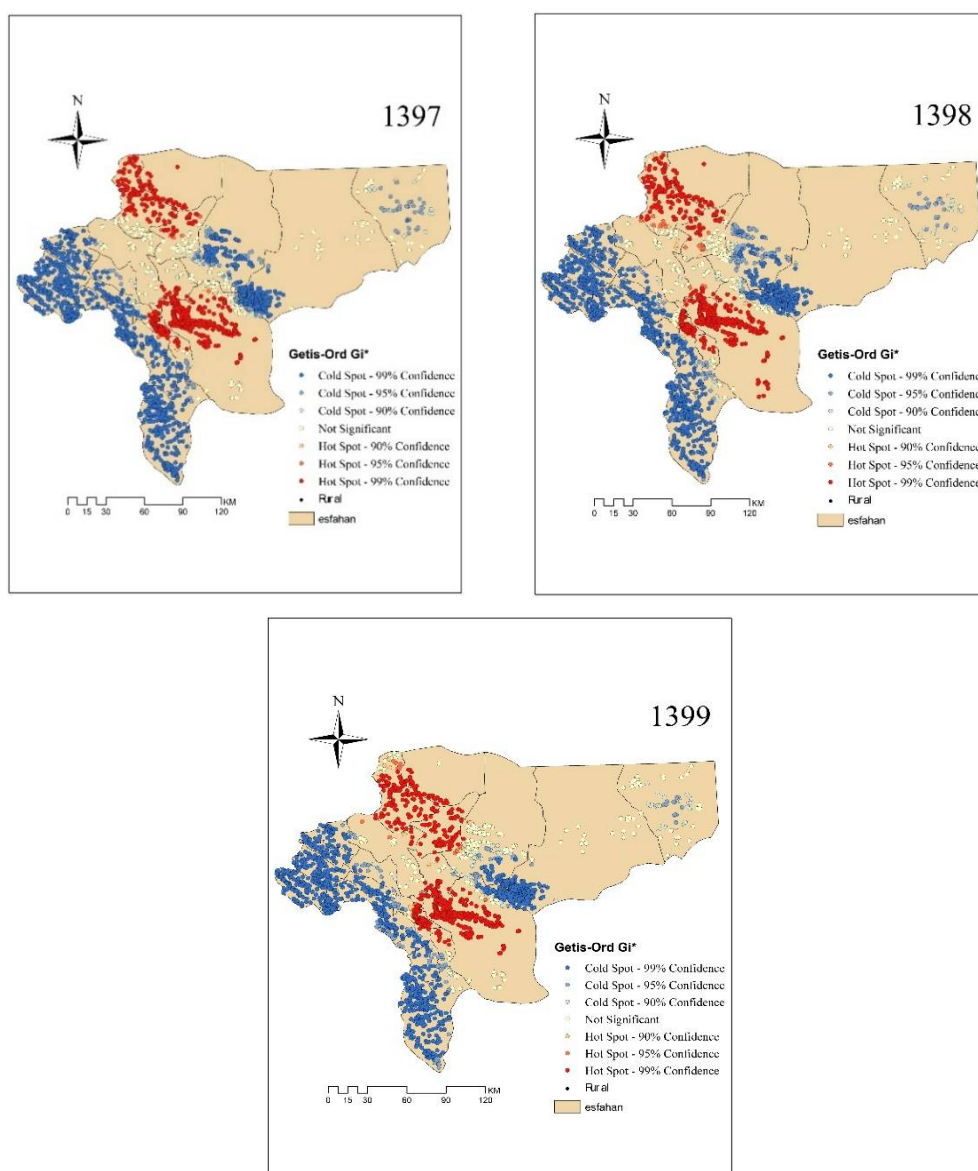
در شکل (5)، با استفاده از شاخص آماری Getis-ord G_i^* شناسایی و تحلیل توزیع بیماری مشاهده می‌شود. بر اساس آخرین گزارش از مرکز آمار ایران، استان اصفهان 114 شهرستان و 1800 روستا است. این تحقیق در محدوده وسیع روستاها انجام شده، بدلیل آنکه بیشتر موارد گزارش حضور پشه ناقل بیماری سالک پستی در روستاهای استان اصفهان رخ داده‌است. همانطور که از شکل (5) مشخص است، روستاهایی که در قسمت‌های شمال غربی، بخش مرکزی و جنوبی استان واقع شده‌اند بیشترین آمار بیماری سالک را نشان داده‌است. همچنین، روستاهای واقع در غرب و بخش‌های جنوب غربی و شرق استان اصفهان، روستاهایی هستند که کمترین آمار را از لحاظ حضور بیماری سالک در میان افراد داشته‌است. به طور تقریبی، توزیع و الگوی روند

بیماری سالک در هر سه سال مطالعه روندی مشخص و یکسان را نشان می‌دهد و همچنین نشان داده‌است که در سال 1398 بیشتر بوده‌است. قسمت‌های جنوبی استان، حضور پشه ناقل بیماری سالک در

قسمت‌های جنوبی استان، حضور پشه ناقل بیماری سالک در سال 1398 بیشتر بوده‌است.

جدول (1): نتایج شاخص جهانی موران

شاخص موران	Z-score	P-value	الگوی توزیع	سال
0/3598	92/33	<0/05	خوشه ای	1397
0/3369	68/58	<0/05	خوشه ای	1398
0/34	68/58	<0/05	خوشه ای	1399



شکل (5): نتایج شاخص Getis-ord Gi*

در شهرستان آران و بیدگل در قسمت شمال شرقی اصفهان، روستاهای وادقان، سار، ون، ارمک، رحق، یحیی آباد و همچنین روستاهای روران، شیدان، عباس آباد، کلودان، هرزم آباد، زغمار، ازوارچه و زیار در شهرستان‌های اصفهان و شهرضا صحرای امیریه و صحرای کلاه قاضی در قسمت مرکزی، در شهرستان

در شهرستان آران و بیدگل در قسمت شمال شرقی اصفهان، روستاهای وادقان، سار، ون، ارمک، رحق، یحیی آباد و همچنین روستاهای روران، شیدان، عباس آباد، کلودان، هرزم آباد، زغمار، ازوارچه و زیار در شهرستان‌های اصفهان و شهرضا صحرای امیریه و صحرای کلاه قاضی در قسمت مرکزی، در شهرستان

فلاورجان چهار برج، قلعه سرخ، سیاه افشار، دارگان، علی شاهدان، سمسار و شهرستان مبارکه روستاهای بروجرد، اکبر آباد، میر آباد، خرم آباد، اسد آباد، جوهرستان به عنوان مناطق داغ بیماری سالک در مرکز استان اصفهان مشاهده می‌شوند. در مقابل، روستاهای گوهان، کمران، ویهه، ننادگان، موغان، آقاگل و مغاندر در قسمت‌های شرقی استان، همچنین روستاهای علی‌آباد، گردمای هواور، پرکستان، برنگان، دورک، دره عرب، حاجی‌آباد، کوشکا، میل درویش، کوشکا در قسمت‌های جنوبی استان و در قسمت شرقی استان روستاهای حسین آباد خان، امیر آباد، توکلیه، جهان آباد، پیوک، بیدچاه در مناطق سرد قرار گرفته‌است.

6-2- بررسی همبستگی میان پارامترهای محیطی و بیماری سالک

پارامترهای محیطی شامل ارتفاع، بارش، دمای هوا، سرعت باد، رطوبت به عنوان پارامترهای موثر در شیوع بیماری سالک در این تحقیق در نظر گرفته شده‌است. با این حال، تاثیر این پارامترها در شیوع بیماری سالک در مناطق مختلف ممکن است متفاوت باشد.

در این تحقیق ارزیابی پارامترهای محیطی موثر بر شیوع بیماری سالک، با استفاده از ضریب همبستگی پیرسون انجام شده‌است. جدول (2) نتایج همبستگی پیرسون را نشان می‌دهد. در این آزمون آماری، فرض صفر بیان می‌کند که بین دو پارامتر رابطه وجود ندارد و در مقابل، فرض یک وجود رابطه بین دو پارامتر را نشان می‌دهد. با توجه به مقادیر سطح معنی‌داری میان پارامترها، هر جا سطح معنی‌داری از مقدار 0/05 کمتر باشد، فرض یک پذیرفته می‌شود و فرض صفر رد می‌شود. میزان همبستگی میان پارامترهای محیطی و شیوع بیماری سالک در سه سال 1397، 1398 و 1399 انجام شد. بر اساس نتایج جدول (2)، پارامترهای سرعت باد، ارتفاع و بارش با شیوع بیماری سالک همبستگی منفی را نشان می‌دهد. بنابراین، با افزایش هر یک از این پارامترها شرایط برای شیوع سالک کاسته می‌شود. در مقابل، پارامترهای محیطی شامل رطوبت و دمای هوا، همبستگی مثبتی را با شیوع سالک نشان می‌دهد. در هر سه سال تحقیق، پارامترهای رطوبت و بارش موثرترین پارامترها بر روی شیوع سالک بوده‌اند، زیرا که بالاترین میزان همبستگی میان پارامترهای محیطی و شیوع سالک را نشان داده‌اند.

جدول (2): نتایج همبستگی پیرسون

پارامترهای محیطی	نرخ بیماری سالک در سال 1397	نرخ بیماری سالک در سال 1398	نرخ بیماری سالک در سال 1399
دمای هوا	+ 0/066	+ 0/056	+ 0/106
بارش	- 0/176	- 0/160	- 0/195
رطوبت	+ 0/154	+ 0/161	+ 0/134
سرعت باد	- 0/095	- 0/088	- 0/107
ارتفاع	- 0/086	- 0/082	- 0/113

روش، یک ناحیه متقارن در اطراف هر نقطه تشکیل داده شده‌است که به ازای هر موقعیت درون ناحیه مورد نظر ارزشی را بر اساس فاصله آن تا نقطه مبدا یا نقاط مبدا در صورت قرار داشتن در محدوده چندین نقطه، ارائه می‌دهد. در نهایت، آنچه در شکل (6) نشان می‌دهد یک سطح پیوسته از برآورد چگالی بیماری سالک می‌باشد.

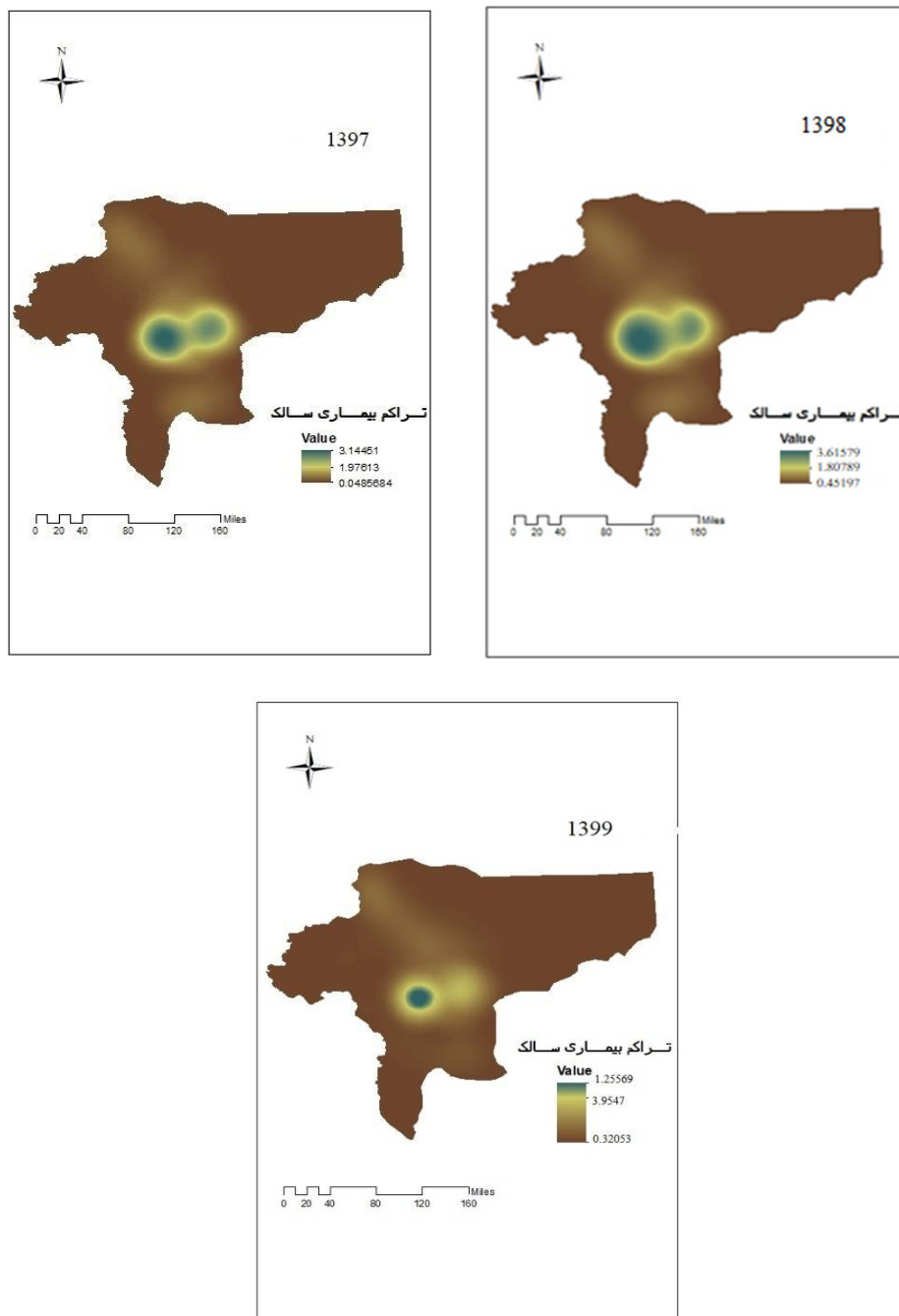
در اینجا، نرخ بیماری سالک مربوط به سال‌های 1397 تا 1399 با استفاده از نرم‌افزار ArcGIS به نقشه رستری

6-3- ارزیابی تراکم بیماری سالک

با استفاده از روش KDE توزیع بیماری سالک انجام شده‌است. این توزیع از طریق پخش کردن احتمال شیوع سالک در یک شعاع معین در اطراف محل وقوع سالک تعریف می‌گردد. در این روش یک محدوده جستجو جهت محاسبه چگالی، بر اساس فاصله فرض شده و محاسبات موردنیاز در این ناحیه جستجو انجام می‌گیرد. در این

در قسمت جنوب استان اصفهان در سال 1398 افزایش کمی در شیوع بیماری سالک مشاهده می‌شود. علاوه بر این، در سال 1399 میزان تعدد بیماری سالک رو به کاهش بوده‌است.

مبنا تبدیل شده‌است. ابعاد واحدهای رستری 30 متر در نظر گرفته شد. نتایج نقشه‌های بدست‌آمده از تراکم سالک نشان داد در سال‌های 1397 و 1398 تغییرات چشم‌گیری در میزان حضور پشه خاکی عامل سالک وجود ندارد و تنها



شکل (6): تراکم بیماری سالک در سال مطالعه

شیوع سالک موثر بوده‌است. همچنین، در سال 1399 تراکم بیماری سالک رو به کاهش رفته‌است و از میزان آن در نواحی جنوب شرقی کمی کاسته شده‌است. بیشترین میزان این بیماری در قسمت مرکزی و کمی در

همانطور که مشخص است، بالاترین تراکم سالک در قسمت‌های جنوب و مرکزی استان اصفهان است. این نواحی از استان اصفهان از ابرفت زاینده رود تشکیل شده‌است و از رطوبت بالایی برخوردار است، که بر روی

جنوب شرقی همچنان در سال 1399 نیز مشاهده می‌شود. نواحی غرب و جنوب غربی استان اصفهان کوهستانی می‌باشد به طوری که بلندترین ارتفاعات در شهرستان‌های فریدون شهر، سمیرم و فریدن قرار دارند. زمستان‌های طولانی بسیار سرد و پربارش از مشخصات بارز این قسمت و قسمت جنوبی استان است. همانطور که از نتایج آزمون همبستگی مشخص شده است، کاهش دمای هوا و افزایش بارش رابطه مستقیمی با کاهش بیماری سالک دارد؛ که به خوبی نتیجه می‌شود پارامترهای دمای هوا و بارش تاثیر زیادی در کاهش سالک در این مناطق روستایی داشته است.

6-4- ارزیابی مدلسازی بیماری سالک

در این تحقیق، مدلسازی بیماری سالک با استفاده از دو الگوریتم قدرتمند یادگیری ماشین شامل شبکه عصبی پرسپترون چند لایه و رگرسیون بردار پشتیبان انجام شد. به منظور مدلسازی بیماری از نقشه‌های پهنه‌بندی تراکم سالک به عنوان پارامتر وابسته و از پارامترهای محیطی تاثیرگذار بر شیوع سالک در استان اصفهان، به عنوان پارامترهای مستقل استفاده گردید. قبل از اجرای مراحل مدلسازی، پیش‌پردازش و نرمالسازی داده‌ها با استفاده از توابع کتابخانه زبان پایتون انجام گرفت. در مرحله مدلسازی، بصورت تصادفی 70 درصد داده‌ها به عنوان داده آموزشی و 30 درصد باقی‌مانده به عنوان داده آزمایشی انتخاب شد. به طور معمول 70 درصد داده‌ها برای آموزش و 30 درصد آن‌ها برای آزمایش تخصیص داده خواهند شد، این فرایند به صورت فرآیند به صورت تصادفی اتفاق خواهند افتاد اما، X هر داده با Y همان داده متناظر باقی خواهند ماند. در مقابل، می‌توان با تنظیم مقدار Random State در هر بار اجرا، داده‌ها را با یک روش یکسانی تقسیم نمود.

یکی از عواملی که تاثیر زیادی روی عملکرد الگوریتم دارد این است که اطلاعات در مرحله آموزش و آزمایش به چه صورتی تقسیم‌بندی شده است. این تقسیم‌بندی اطلاعات تاثیر زیادی بر روی قابلیت تعمیم‌دادن مدل می‌گذارد، زیرا که بخشی از اطلاعات در مرحله آموزش از دست خواهد رفت و آن‌ها برای بخش مربوط به آزمایش شبکه کنار گذاشته خواهند شد. بنابراین، در این تحقیق با هدف به حداقل رساندن میزان این آسیب از روش Grid

Search استفاده شده است [35]. در این روش در هر مرحله، مقادیر مختلف در نظر گرفته شده برای هایپرپارامترها با استفاده از یک توالی در 2000 اپک پردازش و در نهایت پارامترهای بهینه انتخاب شدند.

در الگوریتم MLP در ابتدا پارامترهای اولیه معرفی شدند. این پارامترها شامل توابع انتقال و لایه‌های پنهان می‌باشد. در این تحقیق، توابع انتقال Logistic, Tanh, Sigmoid, Relu در نظر گرفته شدند. همچنین، تعداد زیادی از مقادیر متفاوت نورون‌ها برای لایه‌های پنهان به طور تجربی در نظر گرفته شد و در نهایت بهترین مقادیر این الگوریتم با استفاده از روش Grid Search طبق جدول (3) بدست آمده است.

جدول (3): مقادیر هایپرپارامتر الگوریتم MLP

LearningRate	constant
Alpha	0/0001
Solver	adam
HiddenLayerSizes	(7, 7)
Activation	relu

در الگوریتم SVR برای هایپر پارامترها توابع انتقال شامل (Sigmoid, RBF, Poly, Linear, Precomputed) می‌باشد و همچنین مقدار پارامتر C نیز بین [0/1, 1000] در نظر گرفته شد. علاوه بر این، برای پارامتر Gamma مقادیر Scale, Auto و در نهایت مقدار Epsilon در بازه بین [0/0001, 1] در نظر گرفته شد. در نهایت، با استفاده از روش Grid Search بهینه‌ترین مقادیر طبق جدول (4) بدست آمده است.

جدول (4): مقادیر هایپرپارامتر الگوریتم SVR

gamma	epsilon	C	تابع
scale	0/2	1	poly

عملکرد الگوریتم‌ها با استفاده از شاخص جذر میانگین مربعات خطا طبق رابطه (7) و جذر میانگین مربعات خطای نرمال‌شده بر اساس رابطه (8) بررسی گردید تا نشان دهد مدل ایجاد شده تا چه اندازه توانسته است در پیش‌بینی نمونه‌ها موثر باشد و خوب عمل کند.

نشان داده شد. همچنین، نتایج شاخص‌های آماری نشان داد پارامترهای سرعت باد، ارتفاع و بارش دارای همبستگی منفی با شیوع بیماری سالک هستند، به گونه‌ای که افزایش و یا کاهش هر یک از این پارامترها تاثیر معکوسی بر شیوع سالک خواهد داشت. در مقابل، پارامترهای رطوبت و دمای هوا با شیوع بیماری سالک رابطه مستقیمی دارند، به گونه‌ای که با افزایش و یا کاهش هر یک از این پارامترها روند مستقیم در شیوع بیماری سالک شاهد هستیم. نتایج حاصل از ارزیابی‌های آماری نشان داد، افزایش ارتفاع و وزش بادهای تند در کاهش شیوع سالک در نواحی غربی و جنوب غربی استان موثر بوده‌است به طوری که این مناطق به لحاظ بررسی الگوی مکانی بیماری، در مناطق سرد قرار گرفته‌اند. همچنین، قرار گرفتن نقاط روستایی در نواحی جلگه‌ای و مرطوب مسیر زاینده رود در قسمت‌های مرکزی استان عامل دیگری در افزایش شیوع سالک بوده‌است.

بنابراین، همانطور که نتایج ارزیابی‌ها مشاهده می‌شود، پارامترهای محیطی به کار گرفته‌شده در این تحقیق، بر روی افزایش و یا کاهش شیوع بیماری سالک در نقاط روستایی تاثیر زیادی داشته‌اند. علاوه بر این، از نتایج مربوط به آزمون آماری پیرسون مشاهده شده که در هر سه سال تحقیق، پارامترهای رطوبت و بارش موثرترین پارامترها بر روی شیوع سالک بوده‌اند، زیرا که، از نتایج میان همبستگی پارامترهای محیطی و شاخص‌های آماری، به خوبی می‌توان تاثیر هر یک از پارامترها را بر روند شیوع ملاحظه نمود. در ادامه، در تحقیقات آینده می‌توان نتایج پیش‌بینی الگوریتم‌های یادگیری ماشین را برای نرخ بیماری در سال 1400 مقایسه نمود و نتایج را ارزیابی کرد.

نتایج تحقیق نشان می‌دهد با توجه به شناخت بیماری سالک و بالا رفتن آگاهی جامعه کاهش بیماری را در طی سال‌ها اتفاق خواهد افتاد. همچنین، باید بدانیم پارامترهای محیطی در روند کاهش و افزایش بیماری نقش بسیار زیادی خواهد داشت. به عنوان پیشنهادات در مطالعات آینده می‌توان از پارامترهای محیطی بیشتری و همچنین در مقیاس‌های دیگری همچون مقیاس ماهانه و غیره در مدل‌سازی بیماری سالک استفاده کرد و در کنار آن عواملی مانند مهاجرت، افزایش شهرنشینی و غیره را نیز به عنوان عوامل اجتماعی- اقتصادی مطالعه نمود. در مطالعات آینده همچنین نسبت‌های متفاوتی از مجموعه داده‌های ورودی را به منظور مدل‌سازی بیماری سالک می‌توان در نظر گرفت.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2} \quad (7)$$

$$NRMSE = \frac{RMSE}{O_{ave}} \quad (8)$$

در رابطه (7)، o_i مقدار مشاهده شده، p_i مقدار پیش‌بینی شده و N تعداد داده‌های مشاهده شده است. همچنین، در رابطه (8)، o_{ave} میانگین مقادیر اندازه‌گیری شده می‌باشد [36]. جدول (5) نتایج ارزیابی خطا را بین نتایج مدل و مقدار واقعی تراکم بیماری سالک از میان داده‌های مربوط به سال‌های 1397 تا 1399 را نشان می‌دهد.

جدول (5): نتایج عملکرد الگوریتم‌های یادگیری ماشین

الگوریتم یادگیری ماشین	NRMSE	RMSE
شبکه عصبی پرسپترون چند لایه	11/45	0/348
رگرسیون بردار پشتیبان	12/08	0/170

همانطور که مشخص است، الگوریتم رگرسیون بردار پشتیبان به خوبی توانسته است روند غیر خطی شیوع بیماری سالک را در مقایسه با الگوریتم شبکه عصبی، با توجه به آموزش، که به شبکه داده شده است پیدا کند.

7- نتیجه‌گیری و پیشنهادات

بیماری سالک یک مشکل بهداشت عمومی است که باعث از دست رفتن جان انسان‌ها در سراسر جهان می‌شود. بنابراین، ارزیابی مکانی این بیماری و کاهش اثرات مخرب آن می‌تواند به برنامه‌ریزی و استراتژی‌های طراحی شده برای کنترل این بیماری کمک کند. الگوریتم‌های یادگیری ماشین نیز می‌توانند به عنوان ابزاری برای پیش‌بینی توزیع سالک و شناسایی عوامل تاثیرگذار بر توزیع سالک استفاده شوند. در این مطالعه از الگوریتم‌های رگرسیون بردار پشتیبان و شبکه عصبی پرسپترون چند لایه برای مدل‌سازی بیماری سالک استفاده گردید. پارامترهای محیطی و نرخ بیماری سالک به عنوان پارامترهای مستقل و وابسته در روند مدل‌سازی سالک وارد شدند و در نهایت با استفاده از روش RMSE و NRMSE عملکرد هر یک از الگوریتم‌های رگرسیون بردار پشتیبان و شبکه عصبی پرسپترون چند لایه به ترتیب با مقادیر 0/170، 0/348 و 12/08، 11/45

سپاسگزاری

واگیر بابت در اختیار گذاشتن داده‌های بیماری سالک
مورد استفاده در این تحقیق را دارم.

کمال تشکر را از جناب آقای دکتر محمدرضا
شیرزادی، رئیس مرکز کنترل و پیشگیری بیماری‌های

مراجع

- [1] Carvalho, B. M., et al. (2015). "Ecological niche modelling predicts southward expansion of *Lutzomyia (Nyssomyia) flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania (Leishmania) amazonensis* in South America, under climate change." *PLoS One* 10(11): e0143282.
- [2] Ghatee MA, Haghdoost AA, Kooreshnia F, Kanannejad Z, Parisaie Z, Karamian M, Moshfe A. Role of environmental, climatic risk factors and livestock animals on the occurrence of cutaneous leishmaniasis in newly emerging focus in Iran. *J Infect Public Health*. 2018 May-Jun; 11(3):425-433. doi: 10.1016/j.jiph.2017.12.004. Epub 2017 Dec 26. PMID: 29287805.
- [3] Rajabi, M., et al. (2016). "A spatially explicit agent-based modeling approach for the spread of cutaneous leishmaniasis disease in central Iran, Isfahan." *Environmental modelling & software* 82: 330-346.
- [4] Eslami, S. and Hasanloom, S. (2019). "Modeling the condition of coral reefs using support vector machine regression and applying spectral indices." *Iranian Journal of Marine Technology* 6(1): 31-44.
- [5] Mollalo, A., et al. (2014). "Spatial and spatio-temporal analysis of human brucellosis in Iran." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 108(11): 721-728.
- [6] Mollalo, A., et al. (2014). "Spatial and statistical analyses of the relations between vegetation cover and incidence of cutaneous leishmaniasis in an endemic province, northeast of Iran." *Asian Pacific journal of tropical disease*, 4(3): 176-180.
- [7] Nilfroshzadeh, M.A, Shirani Bidabadi, L., Hosseini, Seyed. M, Fadaei Nobari, R, Jafari, F. (2016). Epidemiology study of cutaneous leishmaniasis in Isfahan province during the years 1380-1390, *Journal of Isfahan Medical School*, 32(315), 2241-2251.
- [8] Murad, A. and B. F. Khashoggi (2020). "Using GIS for disease mapping and clustering in Jeddah, Saudi Arabia." *ISPRS International Journal of Geo-Information*, 9(5): 328.
- [9] Nia, A. M. and A. Alimohammadi (2015). "Geographic Information System-Based Analysis of the Spatial Distribution of Zoonotic Visceral Leishmaniasis and Crimean Congo Haemorrhagic Fever in Provinces of Iran."
- [10] Yu, W., Liu, T., Valdez, R. et al. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 10, 16 (2010). <https://doi.org/10.1186/1472-6947-10-16>.
- [11] AhangarCani, M., et al. (2016). "Predictive Map of Spatio-Temporal Distribution of Leptospirosis Using Geographical Weighted Regression and Multilayer Perceptron Neural Network Methods." *Journal of Geomatics Science and Technology* 6(2): 79-98.
- [12] Akhavan, P., et al. (2014). "Risk Mapping of Cutaneous Leishmaniasis via a Fuzzy C Means-based Neuro-Fuzzy Inference System." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- [13] Shirzadi, M. (2012). "Leishmaniasis skin care guide (leishmaniasis) in Iran." Iran, Ministry of Health, Treatment and Medical Education, Disease Management Center Office of Joint Diseases Control between humans and animals.
- [14] Moore, D. A. and T. E. Carpenter (1999). "Spatial analytical methods and geographic information systems: use in health research and epidemiology." *Epidemiologic reviews*, 21(2): 143-161.
- [15] Pasini, A., et al. (2020). "Neural network modelling for estimating linear and nonlinear influences of meteorological variables on *Sergentomyia minuta* abundance using small datasets." *Ecological Informatics*. 65: 101055.
- [16] Tabasi, M. and A. A. Alesheikh (2017). "A Review of the Applications of Agent Based Simulation in Epidemic Diseases (Case study: Cutaneous Leishmaniasis)." *Geospatial Engineering Journal*, 8(2): 11-23.

- [17] Dehghani, A., et al. (2019). "Epidemiological Study and Spatial Modeling of Cutaneous Leishmaniasis in Bushehr Province Using the Geographic Information System (GIS) from 2011 to 2015." *Journal of Community Health Research*.
- [18] Franke, C. R., et al. (2002). "Trends in the temporal and spatial distribution of visceral and cutaneous leishmaniasis in the state of Bahia, Brazil, from 1985 to 1999." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(3): 236-241.
- [19] Ramos, W. R., et al. (2014). "Anthropic effects on sand fly (Diptera: Psychodidae) abundance and diversity in an Amazonian rural settlement, Brazil." *Acta tropica*, 139: 44-52.
- [20] Mollalo, A., et al. (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*, 62(1): 18-28.
- [21] Mansour, S. (2016). "Spatial analysis of public health facilities in Riyadh Governorate, Saudi Arabia: a GIS-based study to assess geographic variations of service provision and accessibility." *Geo-spatial Information Science*, 19(1): 26-38.
- [22] Liu, G., et al. (2013). "The use of spatial autocorrelation analysis to identify PAHs pollution hotspots at an industrially contaminated site." *Environmental monitoring and assessment*, 185(11): 9549-9558.
- [23] Mathur, M. (2015). "Spatial autocorrelation analysis in plant population: An overview." *Journal of Applied and Natural Science*, 7(1): 501-513.
- [24] Rogers, D. J. and S. E. Randolph (2003). "Studying the global distribution of infectious diseases using GIS and RS." *Nature Reviews Microbiology*, 1(3): 231-237.
- [25] Darand, M., et al. (2017). "Spatial autocorrelation analysis of extreme precipitation in Iran." *Russian Meteorology and Hydrology*, 42(6): 415-424.
- [26] Shafabakhsh, G. A., et al. (2017). "GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran." *Journal of traffic and transportation engineering (English edition)* 4(3): 290-299.
- [27] Węglarczyk, S. (2018). "Kernel density estimation and its application." Volume 23, 2018.
- [28] Ameri, M. and Molayem. M. (2006). "Application of artificial neural networks for analysis of flexible pavements." *International Journal of Industrial Engineering and Production Management, International Journal of Engineering* 17(5).
- [29] Rezaee, Z., et al. (2020). "Application of Time Series Models in Business Research: Correlation, Association, Causation." *Sustainability*, 12(12): 4833.
- [30] Neenwi, S., et al. (2013). "Predicting the Nigerian stock market using artificial neural network." *European Journal of Computer Science and Information* 1(1): 30-39.
- [31] Astray, G., et al. (2010). "The use of artificial neural networks to forecast biological atmospheric allergens or pathogens only as *Alternaria* spores." *Journal of Environmental Monitoring*, 12(11): 2145-2152.
- [32] Bahrami, B. and A. Ghorbani (2015). "Evaluation of Application of Neural Network and Regression Models to Predict Species Diversity Using Some Soil and Physiographic Factors (Case Study: Urmia Ruin Watershed)". *Journal of Natural Ecosystems of Iran*. 5 (2): 65 -80.
- [33] Yu, W., Liu, T., Valdez, R. et al. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 10, 16 (2010). <https://doi.org/10.1186/1472-6947-10-16>
- [34] Huang, K.-Y. And K.-J. Chen (2011). "Multilayer perceptron for prediction of 2006 world cup football game." *Advances in Artificial Neural Systems* 2011.
- [35] Mollalo, A., et al. (2018). "Machine learning approaches in GIS-based ecological modeling of the sand fly *Phlebotomus papatasi*, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran." *Acta tropica* 188: 187-194.
- [36] Abd Latif, Z. and M. H. Mohamad (2015). Mapping of dengue outbreak distribution using spatial statistics and geographical information system. 2015 2nd International Conference on Information Science and Security (ICISS), IEEE.