



مدلسازی مکانی - زمانی شیوع بیماری مalaria

زهره خاکشور^{۱*}، مهدی عربی^۲ و فرهاد حسینعلی^۳

- ۱- دانشجوی کارشناسی ارشد سیستم اطلاعات جغرافیایی دانشگاه تربیت دبیر شهید رجایی
- ۲- استادیار دانشکده نقشه‌برداری دانشگاه تربیت دبیر شهید رجایی
- ۳- استادیار دانشکده نقشه‌برداری دانشگاه تربیت دبیر شهید رجایی

چکیده:

مalaria یک بیماری سازگار با محیط و یکی از بیماری‌های واگیر در نواحی جنوب و جنوب‌شرق ایران است. هر ساله افراد زیادی در اثر ابتلاء آن جان خود را از دست می‌دهند. با گسترش علم و تکنولوژی نیازمند ابزاری نوین جهت مطالعه و بررسی مشکلات بهداشت از جمله بیماری هستیم ازین رو سیستم اطلاعات جغرافیایی (GIS) را می‌توان یکی از این ابزارها برای بررسی و تحلیل بیماری دانست. ایجاد مدلی برای پیش‌بینی میزان شیوع بیماری به کمک آنالیزهای GIS و روش‌های هوش مصنوعی کمک فراوانی در شناخت بیماری می‌کند. در این مطالعه با استفاده از دو روش شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان که هر دو از روش‌های یادگیری با نظرات هستند و با در نظر گرفتن عوامل هواشناسی و محیطی تاثیرگذار بر این بیماری شامل رطوبت نسبی، دما، حداکثر سرعت باد، میزان بارندگی، ارتفاع، پوشش گیاهی، میزان بارندگی، فاصله از آب‌های راکد و رودخانه‌ها و داده‌های مربوط به روش‌های جلوگیری از شیوع malaria و داده‌های مربوط به نرخ وقوع انگل malaria از نوع بیوواکس و فالسیپارام در ۵ سال (۱۳۸۹-۱۳۹۳) به مدل‌سازی میزان شیوع بیماری در نرم‌افزار MATLAB پرداختیم. دو مقدار جذر میانگین مربعات خطأ و ضریب همبستگی بین خروجی شبکه و خروجی مطلوب به عنوان معیاری برای کارایی مدل انتخاب شد، با مقایسه نتایج حاصل از این دو روش، روش رگرسیون بردار پشتیبان با انتخاب مناسب پارامترهای آن دارای جذر میانگین مربعات خطأ کمتر و دارای ضریب همبستگی بیشتر نسبت به شبکه عصبی بهتر توانست میزان شیوع بیماری را مدل و محاسبه کند.

واژه‌های کلیدی: malaria، شیوع، سیستم اطلاعات جغرافیایی، رگرسیون بردار پشتیبان، شبکه عصبی

۱- مقدمه

لازمه‌ی مدیریت صحیح بهداشت عمومی ایجاد زمینه‌های تصمیم‌گیری مناسب بر اساس حجم عظیمی از اطلاعات مکان مرجع موجود در بخش بهداشت و درمان کشور است. در شناخت و بررسی یک بیماری سه عامل اساسی زمان، مکان و شخص بیمار مورد بررسی قرار می‌گیرد.^[۱] اکثر عوامل ایجاد کننده بیماری مکانی هستند به عبارت دیگر توزیع و تمرکز آن در مکان‌های مختلف متفاوت است، دسترسی به این اطلاعات مکانی و تحلیل آن می‌تواند موجبات ارتقای سلامت را فراهم آورد. سازمان‌ها به ابزارهای نیازمندند تا به کنترل و مدیریت بیماری‌ها کمک کند، از آنجایی که مسائل بهداشت عمومی و انتشار بیماری‌ها با جغرافیای منطقه رابطه مستقیم دارد و همواره دارای بعد مکانی هستند، GIS^[۲] می‌تواند نقش بسزایی در مدیریت و برنامه‌ریزی مسائل بهداشت عمومی و مطالعه بیماری‌ها داشته باشد. GIS این امکان را به محققین امور بهداشت می‌دهد تا داده‌ها از منابع مختلف را با هم تلفیق کنند و آنالیزهای مختلف را روی اطلاعات مکانی می‌دهد. [۲] در بررسی بیماری با آندمیک بودن برخی بیماری‌ها مواجه می‌شویم، آندمیک به وجود دائمی یک بیماری در یک منطقه جغرافیایی مشخص گفته می‌شود، جنوب و جنوب‌شرق ایران، شامل بخشی از استان سیستان و بلوچستان یکی از نواحی آندمیک بیماری مalaria است. مalaria یک بیماری انگلی است و عامل بیماری با گزش پشه‌ی ماده آنوفل به انسان منتقل می‌شود. عامل این بیماری، انگل‌های تک سلولی از گونه انگل Malaria^[۳] می‌باشد. [۳] این بیماری در صورت عدم توجه به موقع، می‌تواند نتایج مرگباری را در پی داشته باشد بنا بر گفته سازمان بهداشت جهانی (WHO)^[۴] ایران در مرحله‌ی قبل از حذف کامل بیماری قرار دارد. امروزه مدل‌سازی Malaria و ایجاد سیستم هشدار Malaria فناوری مفید و اقتصادی برای تسهیل کنترل Malaria با استفاده از سیستم اطلاعات جغرافیایی است. [۵] از جمله روش‌هایی که می‌توان برای مدل‌سازی میزان شیوع بیماری استفاده کرد، روش‌های یادگیری ماشینی می‌باشد، در این مطالعه به ارزیابی روش مدل‌سازی روش رگرسیون بردار پشتیبان (SVR)^[۶] که یکی از روش‌های یادگیری نظارت شده می‌باشد پرداختیم و نتایج حاصل از آن را با روش شبکه عصبی مقایسه نمودیم.

۱-۱- مروری بر پیشینه تحقیق

مدل‌سازی بیماری کار بیچیده‌ای است چرا که عوامل و پارامترهای زیادی در ایجاد و شیوع بیماری موثر می‌باشد و امکان بررسی تمام این عوامل و پارامترها وجود ندارد. با این وجود در زمینه مدل‌سازی و تهیه‌ی نقشه‌ی آسیب پذیری بیماری کارهای زیادی انجام شده که به مختصراً از آن‌ها اشاره می‌شود:

حلیمی و همکاران در سال ۲۰۱۴، در این مطالعه ارتباط بین فاکتورهای هواشناسی و بروز انگل Malaria به صورت سالیانه^[۷] (API) در استان سیستان و بلوچستان در جنوب ایران با استفاده از مدل آماری مکانی بیان می‌کند. در این مطالعه از مدل رگرسیون وزن‌دار جغرافیایی (GWR)^[۸] برای پیش‌بینی API با استفاده سه فاکتور میزان بارندگی سالیانه، رطوبت نسبی سالیانه و میانگین دمای سالیانه استفاده شد. در این مطالعه ضریب همبستگی بین مقادیر پیش‌بینی شده و مقادیر واقعی مقدار ۰.۵۱ است. این نشان می‌دهد علاوه بر فاکتورهای هواشناسی، عواملی مثل نوع زندگی، موقعیت‌های اطراف استان در کشور پاکستان و افغانستان نیز باید برای بررسی همه‌گیری^[۹] در نظر گرفته شود. [۷]

^۱-Geographic Information System

^۲-Plasmodium

^۳-World Health Organization

^۴-Support Vector Regression

^۵-Annual Parasite Incidence (API)

^۶-Geographic Weighted Regression (GWR)

^۷-Epidemiology

حنفی و همکاران در سال ۲۰۱۲ به بررسی بیماری مalaria در شهرستان بشاغرد پرداختند. هدف ایشان تعیین موقعیت بیماری و تولید نقشه آسیب‌پذیری بیماری Malaria بوده است. در این مطالعه پارامترهای دما، رطوبت نسبی، ارتفاع، شیب، فاصله از رودخانه و روش‌های کنترل بیماری برای تهیه نقشه آسیب‌پذیر بیماری استفاده شد. ایشان نقاط حساس^۱ بیماری را به روش وزن دهی مبتنی بر شواهد^۲ برای خطر شیوع Malaria به دست آورده‌اند. نقشه آسیب‌پذیری توسط همپوشانی وزندار لایه‌های خطرات زیست محیطی، کاربری اراضی، نرخ وقوع Malaria، پارامترهای توسعه و روش‌های جلوگیری از شیوع تهیه شده است.^[۸]

آذر مهر و همکاران در سال ۱۳۸۷ با طراحی مدلی نحوه انتشار مکانی Malaria را با گذشت زمان شبیه‌سازی کردند. ایشان در این شبیه‌سازی از روش اتوماتیک سلولی و مدل همه‌گیری SIR استفاده نمودند. در این مدل تاثیر هر پارامتر مثل فاصله از آب‌های راکد و تاثیر همزمان دو پارامتر بر چگونگی انتشار مکانی-زمانی بررسی شد. در این شبیه‌سازی از پارامترهای هواشناسی مثل درجه حرارت و رطوبت و پارامترهای محیطی مثل فاصله از آب راکد و جمعیت انسان در نظر گرفته شد. و این لایه‌ها برای ورود به مدل توسط کارشناس وزن دهی و طبقه‌بندی مجدد شدند. در نهایت آلودگی در بازه‌های زمانی مختلف شبیه‌سازی گردید.^[۹]

ارلاندو پی زاکاریاس و بستر در سال ۲۰۱۳، به بررسی مدلی برای پیش‌بینی میزان شیوع Malaria پرداختند، ایشان در این مطالعه فاکتورهایی از جمله داده‌های مربوط به بیماری در ۱۰ سال (۱۹۹۹-۲۰۰۸) مربوط به کشور موزامبیک، داده‌های هواشناسی شامل دما، بارندگی، رطوبت و درصد سهم پاشی ابیابی اماکن داخلی استفاده شد. در این مطالعه دو مدل جنگل‌های تصادفی (RFs) و ماشین بردار پشتیبان^۳ (SVM) برای پیش‌بینی Malaria استفاده شد. در این مطالعه با مقایسه میانگین مربعات خطأ (MSE)^۴ مدلی مناسب برای این امر باشد.^[۱۰]

۲- روش کار

مدل‌های سنتی زیادی برای پیش‌بینی شیوع Malaria وجود دارد، استفاده از روش‌های کلاسیک آماری، نیازمند نمونه‌هایی با حجم بالا و همچنین نیازمند زمان و تلاش برای آنالیز داده‌ها می‌باشد، که ممکن است در عمل فراهم کردن آن میسر نباشد. یکی از روش‌هایی که برای حل چنین مشکلاتی مورد استفاده قرار می‌گیرد، استفاده از الگوریتم‌های یادگیری ماشینی است. مدل‌های محاسباتی با استفاده از تکنیک‌های یادگیری ماشینی توسعه داده شده اند و امروزه برای تشخیص و پیش‌بینی بسیاری از بیماری‌ها مفید هستند.^[۱۱] در مدل‌های محاسباتی باید پارامترهای شیوع Malaria همچنین یادگیری به قدر کافی باشد تا این تکنیک به شکل موثری بتواند، پیش‌بینی را انجام دهد. در مقایسه با روش‌های سنتی، این روش‌ها نیاز به اطلاعات عمیق از وضعیت ندارد. ماشین بردار پشتیبان یک تکنیک جدید از روش‌های یادگیری ماشین است و می‌توان محبوبیت کنونی آن را با محبوبیت شبکه‌های عصبی در دهه‌های گذشته مقایسه کرد.^[۱۲]

۲-۱- ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان، یک تکنیک یادگیری نظارت شده برای مسائل طبقه‌بندی و رگرسیون است که ریشه‌های آن در نظریه‌ی یادگیری آماری و اپنیک^۵ در سال ۱۹۹۵ است.^[۱۳] برای آموزش داده‌ها SVM از یک ابر صفحه چند بعدی استفاده می‌کند تا اختلاف بین دو کلاس را با حداقل کردن حاشیه بین دو کلاس به صورت بهینه به دست بیاورد. در بیشتر موارد نمی‌توان از یک جداگانه خطی استفاده کرد. یک راه حل برای این مشکل نگاشت داده‌ها به یک

¹ - Hot Spot

²- Evidence Based Weighted Method

³- Support Vector Machine

⁴- Mean Square Error

⁵ - Vapnik

فضا با ابعاد بزرگتر با استفاده از یک تبدیل غیر خطی می‌باشد، به طوریکه دو کلاس در فضای جدید قابل جدا شدن باشد.^[۱۵] تبدیلی که برای این نگاشت استفاده می‌شود،تابع هسته^۱ نامیده می‌شود. از جمله توابع کرنل شناخته شده، کرنل چندجمله‌ای، کرنل گوسی، کرنل سیگموئید و کرنل توابع پایه شعاعی می‌باشد. بر خلاف شبکه‌های عصبی که سعی می‌کند به تعریف پیچیدگی توابع از فضای ورودی بپردازد، ماشین بردار پشتیبان داده‌های غیرخطی را به فضایی با ابعاد بالاتر می‌برد، سپس از توابع خطی ساده برای ایجاد مرزهای جدا کننده خطی در فضای جدید استفاده می‌کند در ماشین های بردار پشتیبان برخلاف شبکه‌های عصبی و رگرسیون های خطی که به وسیله مینیمم کردن قدر مطلق خطا یا توان دوم خطا عمل می‌کنند، آنها ریسک عملیاتی را مینیمم می‌کنند.^[۱۶ و ۱۷] ماشین بردار پشتیبان در پیش‌بینی‌ها نسبت به شبکه‌های عصبی از دقت بالاتری برخوردار است از طرفی تعیین وضعیت شبکه و قابلیت تعمیم شبکه عصبی برای وظایف مدل‌سازی/یادگیری هنوز به خوبی حل نشده است، در حالی که مدل ماشین بردار پشتیبان به خوبی تعمیم‌پذیراست همچنین با استفاده از کرنل‌های غیر خطی قادر به تصمیم‌گیری غیر خطی نیز می‌باشد. انتخاب کرنل مناسب برای ماشین بردار پشتیبان، منجر به برتری آن نسبت به سایر رویکردهای مبتنی بر تصمیم‌گیری خطی شده است.^[۱۸]

۱-۱-۲- رگرسیون بردار پشتیبان (SVR)

مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده (الف) مدل طبقه‌بندی ماشین بردار پشتیبان و (ب) مدل بردار پشتیبان رگرسیون تقسیم بندی می‌شوند. از مدل طبقه‌بندی ماشین بردار پشتیبان جهت حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند استفاده می‌گردد و مدل رگرسیون ماشین بردار پشتیبان در حل مسائل

پیش‌بینی کاربرد دارد. همانگونه که بیان شد، ماشین بردار پشتیبان مبتنی بر مینیمم کردن ریسک ساختاری^۲ می‌باشد که از تئوری آموزش آماری گرفته شده است.^[۱۹] تابع زیر را در نظر می‌گیریم.

$$f(x) = w^T \cdot x + b \quad (1)$$

که در آن x ورودی، w بردار وزن و b دوره بایاس می‌باشد. مدل رگرسیون در فرمول (۱) با استفاده از تابع نگاشت $\phi(x)$ می‌تواند به صورت زیر در نظر گرفته شود.^[۲۰]

$$f(x) = w^T \cdot \phi(x) + b \quad (2)$$

با استفاده از رابطه (۲) داده‌های ورودی اصلی به فضا با ابعاد بالاتر برده می‌شود که در آن مسئله تفکیک‌پذیر غیر خطی در همین فضای جدید به تفکیک‌پذیر خطی تبدیل می‌شود، تابع نگاشت متاثر از بردار وزن w و بایاس b است. مقادیر مناسب این دو پارامتر از حداقل کردن خطای تخمین و یکسان در نظر گرفتن وزن‌ها به دست می‌آید. در واقع هدف اصلی کمینه کردن تابع زیر می‌باشد.^[۲۰]

$$R(f) = \frac{1}{2} w^T w + C.E(x, y, f) \quad (3)$$

E در رابطه بالا تابع تلفات می‌باشد و C پارامتری است که بر حسب حداقل کردن خطای تخمین و یکسان در نظر گرفتن وزن‌ها تنظیم می‌شود. فرمول (۳) را می‌توان به شکل زیر نوشت.

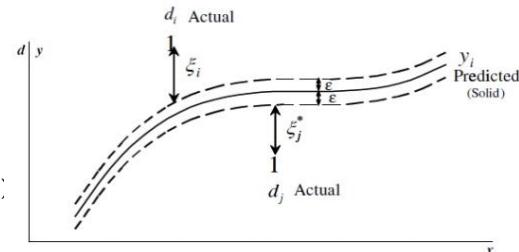
¹ -Kernel

² -Structural Risk Minimization (SRM)

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^N (\xi_i - \xi_i^*) \quad \text{رابطه (4)}$$

قیود:

$$\begin{aligned} d_i - w^T \phi(x_i) - b &\leq \varepsilon + \xi_i & i = 1, 2, \dots, N \\ w^T \phi(x_i) - d_i + b &\leq \varepsilon + \xi_i^* & i = 1, 2, \dots, N \\ \xi_i, \xi_i^* &\geq 0 & i = 1, 2, \dots, N \end{aligned} \quad (5)$$



شکل 1: پارامترهای مورد استفاده در ماشین بردار

پشتیبان

در رابطه های (4) و (5) و شکل (1) x_i ورودی که با استفاده از تابع نگاشت ϕ به فضای ابعاد بالاتر برده می شود، d_i خروجی، C هزینه خطای ξ_i و خطای ξ_i^* پایین تر و ξ_i خطای یادگیری بالاتر است. درینجا ξ_i و ξ_i^* متغیرهای کمکی برای قیود اجرا نشدنی در مسئله بهینه سازی می باشند. برای حل مسئله بهینه سازی می توان از روش لAGRANZ استفاده نمود و تابع لAGRANZ را مطابق زیر نوشت.

$$\begin{aligned} L = R(f) - \sum_{i=1}^N (\eta_i \xi_i - \eta_i^* \xi_i^*) - \sum_{i=1}^N \alpha_i (\varepsilon + \xi_i - d_i + w^T \phi(x_i) + b) \\ - \sum_{i=1}^N \alpha_i^* (\varepsilon + \xi_i + d_i + w^T \phi(x_i) - b) \end{aligned} \quad \text{رابطه (6)}$$

در رابطه (6) مقادیر η_i ، η_i^* ، α_i و α_i^* ضرایب لAGRANZ هستند. مشتق L نسبت به پارامترهای w ، ξ_i و ξ_i^* را برابر صفر قرار می دهیم در نتیجه مسئله بهینه سازی به شکل زیر در می آید:

$$\max -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \phi_i(x) \phi_j(x) - \varepsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) + \sum_{i=1}^N d_i (\alpha_i - \alpha_i^*) \quad \text{رابطه (7)}$$

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \quad \alpha_i, \alpha_i^* \in [0, C] \quad \text{رابطه (8)}$$

با توجه به رابطه (7) و (8) لازم است $K(x_i, x_j) = \phi_i(x) \phi_j(x)$ که کرنل نامیده می شود، مشخص باشد، می توان نوشت.

$$w = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \phi(x_i) \quad \text{رابطه (9)}$$

$$w = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad \text{رابطه (10)}$$

برای رسیدن مسئله به جواب بهینه باید در شرایطی به نام شرایط کوهن تاکر(KKT)^۱ صدق کند که با برقراری این شرایط مقدار مقدار پارامتر b دست می‌آید.

$$\max\{-\varepsilon + d_i + w^T \cdot \phi(x_i) | \alpha_i < C \text{ or } \alpha_i^* > 0\} \leq b \leq \min\{-\varepsilon + d_i + w^T \cdot \phi(x_i) | \alpha_i > 0 \text{ or } \alpha_i^* < C\} \quad (11)$$

در رابطه (11) به طور متداول ازتابع گوسی به صورت $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\delta^2)$ استفاده می‌شود که در این تحقیق نیز از این کرنل استفاده شده است. که δ پارامتر کرنل می‌باشد. در ماشین بردار رگرسیون باید سه پارامتر C ، ε و پارامتر کرنل تعیین شود که نقش مهمی در پیش‌بینی توسط ماشین بردار رگرسیون دارد. پارامتر C درجه ریسک عملیاتی SVR می‌باشد.^۲ عرض بازه خطرا مشخص می‌کند که در شکل (۱) نشان داده شده است. δ عرض تابع گوسی از تابع کرنل است. [۲۰]

۲- شبکه عصبی

شبکه‌های عصبی مصنوعی یک خانواده از مدل‌ها با الهام از شبکه عصبی بیولوژیکی(شبکه عصبی حیوانات به ویژه مغز) است که برای برآورد و توابع تقریبی که وابستگی به تعداد زیادی از ورودی‌ها یا عوامل ناشناخته دارد، استفاده می‌شود. شبکه‌های عصبی مصنوعی به طور کلی سیستمی از نرون‌های متصل به هم است که پیام‌هایی را بین یکدیگر تبادل می‌کنند. ارتباط بین نورون‌ها می‌تواند دارای یک وزن عددی می‌باشد که بر اساس تجربه تنظیم شده است و همچنین با ایجاد یک شبکه عصبی طبیعی، امکان یادگیری برای شبکه فراهم می‌شود در نتیجه می‌تواند وزن‌ها را تنظیم کند. یک شبکه عصبی معمولاً توسط سه پارامتر تعریف می‌شود:

- ۱- الگوی اتصال بین لایه‌های مختلف از نرون‌ها
- ۲- فرآیند یادگیری برای بهروزسازی وزن اتصالات
- ۳- تابع فعالسازی که مقادیر خروجی یک نرون مصنوعی را بر اساس مقادیر ورودی‌اش تعیین می‌کند.

$$f(x) = K(\sum_i w_i g_i(x)) \quad (12)$$

رابطه (12) به بیان ریاضی تابع شبکه عصبی $f(x)$ به صورت ترکیبی از توابع دیگر (g_i) تعریف می‌شود معمولاً از مجموع وزن دار غیرخطی در شبکه عصبی استفاده می‌شود. که در آن K (ممولاً به تابع فعالسازی بر می‌گردد)، می‌تواند تابع تانزانت هیپربولیک، سیگموئید، خطی، تابع پله‌ای و گوسی باشد [۲۱].

۳- پیاده‌سازی

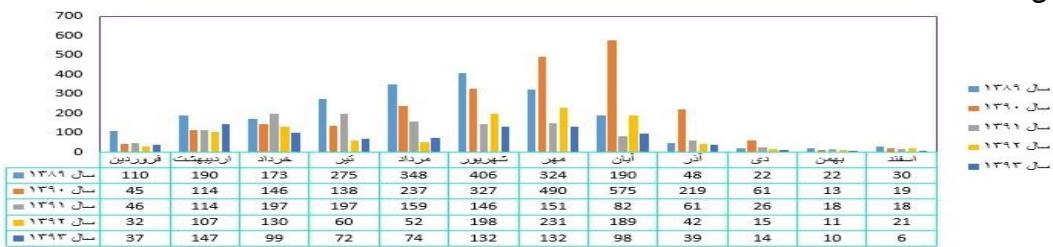
منطقه مورد مطالعه در این پژوهش استان سیستان و بلوچستان که در مختصات جغرافیایی ۲۵ درجه و ۳ دقیقه تا ۳۱ درجه و ۲۷ دقیقه عرض شمالی از خط استوا و ۵۸ درجه و ۵۰ دقیقه تا ۶۳ درجه و ۲۱ دقیقه از نصف‌النهار گرینویچ واقع شده است. استان سیستان و بلوچستان با توجه به موقعیت جغرافیایی خود، از یک طرف تحت تأثیر جريان‌های جوی متعدد مانند "جريان بادی شبکه قاره هند" و به تبع آن باران‌های موسمی اقيانوس هند است و از طرف دیگر تحت تأثیر فشار زیاد عرض‌های متوسط قرار دارد. در این استان با توجه به نزدیکی به مدار راس السرطان آب و هوای بسیار خشک و گرمای شدید مهمن ترین پدیده مشهود اقلیمی آن است

^۱ - Karush Kuhn Tucker (KKT)

۱-۳-داده‌های مورد استفاده

با توجه به پیشینه تحقیق پارامترهای هواشناسی دما، رطوبت نسبی، میزان باندگی، حداکثر سرعت باد و پارامترهای محیطی شامل ارتفاع، پوشش گیاهی، فاصله از آب‌های راکد و رودخانه‌ها و روش‌های مورد استفاده برای جلوگیری از بیماری به عنوان ورودی برای مدل در نظر گرفته شد. علاوه بر پارامترهای تحقیقات قبلی سرعت باد نیز در نظر گرفته شد. این پارامتر نیز به طور غیر مستقیم می‌توان در مدل‌سازی موثر واقع شود [۲۲]

داده‌های ماهانه نرخ وقوع بیماری مalaria در طی ۵ سال (۱۳۸۹-۱۳۹۳) که مطابق گزارشات وزارت بهداشت درمان و آموزش پزشکی کشور می‌باشد. در شکل (۲) نرخ وقوع بیماری در ۵ سال در استان سیستان و بلوچستان نمایش داده شد.



شکل ۲: نرخ وقوع بیماری به صورت ماه به ماه در سال‌های ۱۳۸۹-۱۳۹۳

لایه ارتفاعی منطقه^۱ (DEM)، که از سایت USGS^۲ تهیه شد و با استفاده از آنالیز آمار ناحیه‌ای^۳ متوسط ارتفاع برای هر شهرستان محاسبه شد. لایه شاخص پوشش گیاهی: شاخص پوشش گیاهی نرمال شده (NDVI)^۴ از تصاویر ماهواره‌ای MODIS با پیکسل سایز ۲۵۰ متر تهیه گردید و با استفاده از آنالیز Zonal Statistic^۵ متوسط اساخته پوشش گیاهی برای هر شهرستان محاسبه شد. لایه فاصله از آب‌های راکد و رودخانه^۶: با استفاده از آنالیز فاصله^۷ در امتداد خط مستقیم تهیه شد. با توجه به اینکه در مناطق نزدیک این فاکتور احتمال شیوع بیماری بیشتر است، برای آن یک طبقه‌بندی مجدد^۸ صورت گرفت. لایه مربوط روش‌های جلوگیری از شیوع بیماری: این اطلاعات در قالب فایل اکسل از وزارت بهداشت و درمان آموزش پزشکی کشور تهیه شد. در این پژوهش دو نوع روش جلوگیری لاروکشی^۹ و سم پاشی ابیایی اماكن داخلی^{۱۰} در نظر گرفته شد، در هر شهرستان که از این فاکتورها کمتر بهره‌مند شده باشد امکان شیوع بیماری در آن بیشتر است و در واقع برای آن یک طبقه‌بندی مجدد صورت گرفت. داده‌های هواشناسی: برای بررسی اثر پارامترهای آب و هوایی بر شیوع بیماری Malaria از داده‌های ایستگاه‌های سینوپتیک استان سیستان و بلوچستان، خراسان جنوبی، کرمان، هرمزگان که مربوط داده‌های ماهانه سال ۱۳۸۹ تا ۱۳۹۳ می‌باشد و از سازمان هواشناسی کشور تهیه شد، استفاده گردید. پارامترهای هواشناسی شامل متوسط دمای هوا (درجه سانتی گراد)، متوسط رطوبت هوا، میزان بارش (میلی‌متر)، متوسط حداکثر سرعت باد (متر بر ثانیه) می‌باشد. برای استفاده از هر کدام از پارامترها ابتدا باید لایه‌های اطلاعاتی مربوط به آنها را ایجاد کرد. با توجه به اینکه داده‌های هواشناسی به صورت نقطه‌ای است و نیز به دلیل نبودن ایستگاه‌های سینوپتیک در همه شهرستان‌ها و یا اینکه در بعضی از شهرستان‌ها چند ایستگاه وجود دارد، لازم بود تا از روش‌های درونیابی استفاده شود تا اطلاعات هواشناسی برای کلیه شهرستان‌ها در دسترس باشد. برای هر کدام از پارامترهای هواشناسی میانگین ماهانه آن در ایستگاه‌های

¹-Digital Elevation Model

²- United States Geological Survey

³-Zonal Statistic

⁴- Normalized Difference Vegetation Index

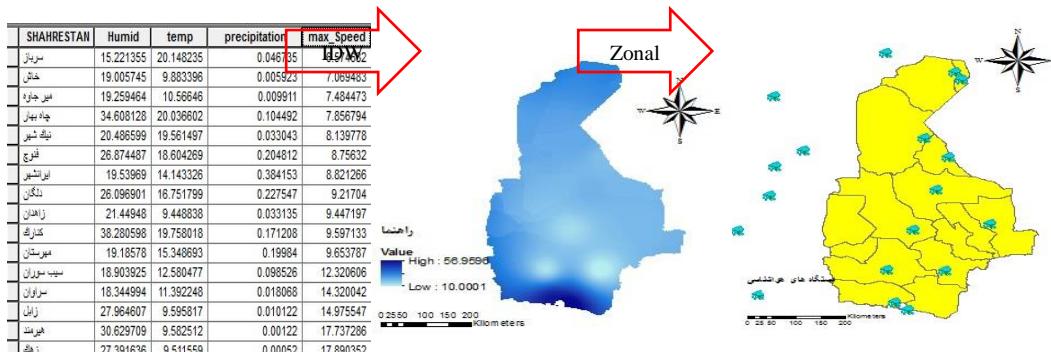
⁵-Distance

⁶-Reclassify

⁷-Larvicide

⁸-Indoor Residual Spraying

سینوپتیک استخراج شد سپس دورنیابی به روش وزن دهنده معکوس فاصله^۱ (IDW) انجام شد و سپس آنالیز Zonal Statistic انجام شد تا میانگین هر پارامتر برای هر شهرستان محاسبه شود. در شکل (۳) این مراحل را به صورت نقشه می بینیم.



شکل ۳: از چپ به راست به ترتیب، اطلاعات ایستگاه‌های هواشناسی به صورت نقطه‌ای، نقشه رستری تولید شده به روش IDW و تبدیل نقشه رستری به جدول و محاسبه‌ی پارامترهای هواشناسی هر شهرستان با استفاده از تحلیل Zonal Statistic

۲-۳- ایجاد مدلی برای پیش‌بینی میزان شیوع مالاریا

هدف از این پژوهش ایجاد مدلی برای پیش‌بینی میزان شیوع بیماری تحت فاکتورهای ذکر شده می‌باشد که از دو روش شبکه عصبی و بردار پشتیبان رگرسیون در محیط برنامه‌نویسی MATLAB استفاده شد. شبکه عصبی استفاده شده در این مطالعه از نوع پرسپترون چند لایه (MLP)^۲ با دو لایه یه پنهان است، ساختار پرسپترون چند لایه به توابع غیر خطی برای فعالسازی نیاز دارد که به طور پیوسته مشتق‌پذیر باشد، توابع فعالسازی زیادی این شرایط را دارا هستند، اما تابع سیگموئید به دلیل سادگی مشتق آن بیشترین کاربرد را دارد. [۲۱] تابع فعالسازی سیگموئید که در لایه‌ی اول دارای ۸ نرون و در لایه‌ی دوم دارای ۲۵ نرون می‌باشد. برای آموزش شبکه، از الگوریتم آموزش لونبرگ-مرکوات استفاده شد، این الگوریتم بگونه‌ای عمل می‌کنند که احتمال در دام افتادن شبکه در یک نقطه حداقل محلی کاهش می‌یابد و همچنین در این الگوریتم آموزش از سیکل‌های آموزشی کمتری استفاده می‌کند و به خطای کمتری می‌رسد. [۲۱]

در روش مدل‌سازی رگرسیون بردار پشتیبان، کرنل گوسی بدلیل عملکرد بهتر آن در پیش‌بینی استفاده گردید. [۲۳] تعیین درست پارامترهای رگرسیون بردار پشتیبان نقش مهمی در نتایج پیش‌بینی دارد، ازین رو پارامترهای آن با سعی و خطا و بررسی نتایج حاصل با نتایج قبلی انتخاب گردید. عرض بازه(^۳) خطا مقدار ۰.۰۵، پارامتر درجه ریسک عملیاتی (C) مقدار ۲۰ و برای پارامتر کرنل (σ) عدد ۴ در نظر گرفته شد.

۲-۴- نتایج دو روش

جدول (۱)، دو مقدار ضریب همبستگی (R) و جذر میانگین مربعات خطای (RMSE)^۴ بین نتایج مدل و مقدار واقعی میزان شیوع بیماری در داده‌های تست در دو روش مدل‌سازی را نشان می‌دهد.

جدول (۱) مقایسه نتایج شبکه عصبی و بردار پشتیبان رگرسیون برای داده‌های تست

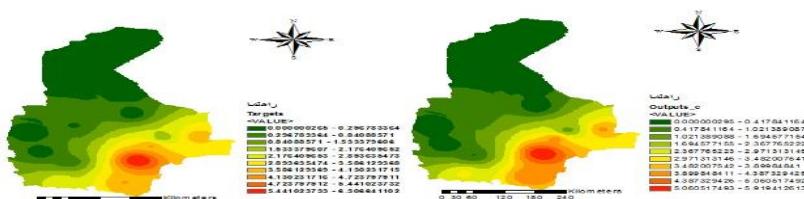
روش	RMSE	R
رگرسیون بردار پشتیبان	۰.۷۸	۰.۹
شبکه عصبی	۱.۲۱	۰.۸۱

^۱-Inverse Distance Weighted

^۲-Multilayer Perceptron

^۳-Root Mean Square Error

مطابق جدول (۱): رگرسیون بردار پشتیبان با میزان خطای کمتر نسبت به شبکه عصبی کاربی بیشتری دارد، لذا با استفاده از این روش نقشه میزان شیوع بیماری در شهریور ماه سال ۱۳۹۳ با استفاده از آنالیز درونیابی IDW تهیه شد و در کنار آن نقشه بیماری در این ماه با استفاده از داده‌های واقعی نیز تهیه گردید. در شکل (۴) می‌توان این دو نقشه را دید. با توجه به دو نقشه، نتایج پیش‌بینی تقریباً به مقادیر واقعی آن نزدیک است.



- [6] M. Brati, H. Keshavarz_valian, M. Habibi-nokhandan, A. Raeisi, L. Faraji & A. salehi-moghadam, Spatial outline of malaria transmission in Iran, Asian Pacific Journal of Tropical Medicine, 2012, 789-795
- [7] M. Halimi, M. Farajzadeh, M. Delavari, A. Takhtardeshir , A. Moradi, Modelling Spatial relationship between climate conditions and annual parasite incidence of malaria in southern part of Sistan&Bloucheatan province of iran using spatial statistic models, Asian pac J Trop Dis 2014 ; 4(supple1):S167-S172
- [8] A.A. Hanafi-Bojd, H. Vatandoost, M.A. Oshaghi, Z. Charrahy, A.A. Haghdoost, G. Zamani, F. M.M. AbediSedaghat , M. Soltani, M. Shahi, A.Raeisi, Spatial analysis and mapping of malaria risk in an endemic area, south of Iran: A GIS based decision making for planning of control, Journal of Acta Tropica,2012, 132-137
- [۹] آذرمهر، م.، مسگری، م.، کریمی، م.، مدل سازی مکانی-زمانی بیماری مalaria با استفاده از Geographic و روش Cellular Automata Information System و فصلنامه بیماری های عفونی و گرمیسری وابسته به انجمن متخصصین بیماری های عفونی و گرمیسری، سال پانزدهم، شماره ۴۸، صفحات ۶۱-۶۹.
- [10] Orlando P. Zacarias, H. Bostr , Comparing Support Vector Regression and Random Forests for Predicting Malaria Incidence in Mozambique, 2013, International Conference on Advances in ICT for Emerging Regions (ICTer): 217 – 221
- [11] V.S.Sriram Tarigoppula, M. Venkateswara Rao,G.V. Satya Narayana, D.S.V.G.K. Kaladhar, T. Pandu RangaVital, 2013,”Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms”, International Journal of Engineering and Innovative Technology (IJEIT) ,Volume 3, Issue 3, September 2013.
- [12] V. Sharma, A. Kumar, L. Panat, G. Karajkhede, A. Lele, Malaria Outbreak Prediction Model Using Machine Learning, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 12, December 2015
- [13] V.N. Vapnik , The Nature of Statistical Learning Theory. Springer Verlag, (1995).
- [14] C. Cortes, and V. Vapnik, Support-vector networks, Machine Learning, vol. 20, pp. 273–297, 1995.
- [15] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, 2nd ed. John Wiley, 2001.
- [16] W.C. Hong, “Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm”, Neurocomputing, Vol. 74, 2096-2107, 2011.
- [17] J. Yin ,“LogP prediction for blocked tripeptides with amino acids descriptors (HMLP) by multiple linear regression and support vector regression”, Procedia Environmental Sciences 8, pp. 173–178, 2011.
- [18] K. Vojislav, Learning and soft computing—Support vector machines, Neural Networks and Fuzzy Logic Models. Massachusetts: The MIT Press; 2001.
- [19] V.N. Vapnik, Statistical learning theory, Wiley, New York, 1998
- [20] C.C. Chang, and C.J. Lin, LIBSVM: A Library for Support Vector Machines. [Online]. Available: (2001)
- [۲۱] سلطانی، سعید، سروش سرداری، مژگان شیخپور و صفری موسوی. ۱۳۸۹. شبکه های عصبی مصنوعی. تهران. موسسه علمی فرهنگی «نص»
- [22] P. Reiter , Climate change and mosquito -borne disease. EnvironHealth Perspect 2001;109 Suppl 1:141-61.
- [23] V. Cherkassky and Y. Ma, “Practical selection of SVM parameters and noise estimation for SVM regression”, Neural Networks, Vol. 17, pp. 113–126, 2004.