

مدلسازی مکانی رخداد بیماری لمپی اسکین در ایران مبتنی بر داده‌های اقلیمی

فاطمه علی بیگی^۱، علیرضا شکیبای*^۲، بابک میرباقری^۳، علی اکبر متکان^۴

۱. دانشجوی کارشناسی ارشد مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی
۲. دانشیار مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی
۳. استادیار مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی
۴. استاد مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی

چکیده

در سال‌های اخیر نگرانی‌ها درباره تاثیر تغییر اقلیم بر تغییر الگوی مکانی و زمانی بیماری‌های واگیر و شرایط زیستگاه گونه‌ها افزایش یافته است. بیماری لمپی اسکین یک بیماری ویروسی با پیامدهای اقتصادی گسترده است که از طریق ویروس کاپری پاکس در گاو ایجاد می‌شود. هدف از این مطالعه، پیش بینی احتمال وقوع بیماری ویروسی لمپی اسکین در جمعیت گاوها برای سال‌های ۲۰۱۴ تا ۲۰۱۸ در ایران است. در این تحقیق از مدل حداکثر آنتروپی (MaxEnt) برای توصیف آشیان اکولوژیکی اقلیمی بیماری استفاده شده است. بدین منظور ابتدا از بین ۱۹ متغیر زیست اقلیمی موجود، ۷ متغیر (۴ متغیر دمایی و ۳ متغیر بارشی) غیرهمبسته و با بیشترین درصد اهمیت نسبی در آزمون جک نایف به همراه متغیرهای ارتفاع و پوشش گیاهی برای مدلسازی به کار گرفته شده است. برای ارزیابی دقت مدل نیز از محاسبه مساحت زیر منحنی ROC استفاده گردید. نقشه احتمال رخداد این بیماری نشان داد که در بخش‌های غرب، شمال غرب، شمال شرق ایران و جنوب دریای مازندران بیشترین احتمال وقوع بیماری لمپی اسکین را دارا بوده و دقت مدل نیز بر اساس شاخص مساحت زیر منحنی ROC برابر ۰.۸۸ می‌باشد. نتایج آزمون جک نایف نشان داد که متغیرهای پیش بینی‌کننده شاخص پوشش گیاهی و بارش سالانه به ترتیب بیشترین اهمیت را در مدل MaxEnt دارا می‌باشند.

کلید واژه‌ها: بیماری لمپی اسکین، مدل حداکثر آنتروپی، داده‌های زیست اقلیمی، سامانه اطلاعات مکانی، ایران

مقدمه

بیماری پوستی لمپی اسکین^۱ یک بیماری حاد ویروسی است که توسط بندپایان خونخوار^۲ انتقال پیدا می‌کند [۱]. این بیماری پوستی گاوی برای اولین بار به عنوان "urticaria pseudo" در سال ۱۹۲۹ در زامبیا معرفی شد [۲]. به طوری که مرزهای خود را به مناطق اکولوژیکی و اقلیمی مختلف مانند آفریقا، آسیا، خاورمیانه و اروپا گسترش داده و اخیراً که از آفریقا وارد کشورهای خاورمیانه شده، بسیار واگیردار است [۳-۵]. براساس گزارش سازمان دامپزشکی کشور این بیماری برای اولین بار در استان کردستان شناسایی، که در سال ۲۰۱۴ به سازمان بین‌المللی بهداشت حیوانات^۳ گزارش شد، بعد از آن از طریق استان‌های شمالی و شمال‌غربی در سراسر ایران به سرعت گسترش یافت [۶]. توسعه فن‌آوری و زمینه‌های کاربردی جدید در GIS و تکنولوژی جمع‌آوری داده‌ها همچون GPS، سنجش از دور و غیره پژوهش را در زمینه مسائل بهداشت و درمان میسر نموده. و باعث ایجاد مطالعات گسترده‌ای در سراسر جهان در مورد بیماری‌ها و مسائل مربوط به آنها به دنبال گشته است [۷].

همانطور که اشاره شد روش اصلی انتقال این بیماری از طریق حشرات می‌باشد. توزیع این حشرات به شدت به اقلیم، شرایط محیطی، فصل، دما، رطوبت نسبی و پوشش گیاهی وابسته است. لذا همین عوامل می‌تواند علت تفاوت در شیوع و مرگ و میر ناشی از این بیماری در مناطق مختلف باشد [۶]. بر این اساس نیاز است این تغییرات موثر بر میزان وقوع این بیماری مدلسازی شود. گرچه بیماری‌های واگیر با روش‌های مختلف ریاضی، مدل‌سازی آماری و تکنیک‌های شبیه‌سازی مکانی^۴ بررسی می‌شود اما کمتر به مدلسازی آشیان بوم‌شناختی^۵ این بیماری و پتانسیل پاسخ این بیماری به تغییرات اقلیمی پرداخته شده است [۸]. مدلسازی آشیان بوم‌شناختی یا مدلسازی توزیع گونه‌ها

^۱ بین موارد حضورگونه در محل (محل وقوع بیماری) با شرایط محیطی ارتباط برقرار می‌کند. در دو دهه اخیر پیشرفت زیادی در مدل‌سازی توزیع گونه‌ها صورت گرفته و به طور گسترده در جغرافیای زیستی، اکولوژی، اپیدمیولوژی، زیست‌شناسی تکامل و محیط زیست به کار می‌رود. مدلسازی آشیان بوم‌شناختی با استفاده از مدل حداکثر آنتروپی^۶ برای پیش‌بینی توزیع احتمال وقوع بیماری به کار گرفته می‌شود [۹-۱۰]. MaxEnt یک الگوریتم یادگیری ماشین چند منظوره است که دقت پیش‌بینی بالایی دارد و توزیع احتمال وقوع بیماری را بر اساس مجموعه داده‌های "فقط-حضور"^۸ و محدودیت‌های محیطی و با تعداد محدود نمونه‌ها تخمین می‌زند. با توجه به این که در اکثر مناطق داده‌های حضور و عدم حضور محدود هستند و اکثراً به صورت داده‌های "فقط-حضور" ثبت شده است. بنابراین استفاده از روش‌هایی که با داده‌های "فقط-حضور" کار می‌کنند ضروری است [۱۱-۱۲].

مواد و روش‌ها

منطقه مورد مطالعه

محدوده مورد مطالعه در این تحقیق کل محدوده کشور ایران می‌باشد. ایران کشوری پهناور در کمربند خشک و نیمه خشک جهان و جنوب غربی آسیا (خاورمیانه) واقع در مرز بین دریای خزر و خلیج فارس با موقعیت جغرافیایی ۲۵.۰۳ تا ۳۹.۴۷ درجه عرض شمالی و ۴۴.۱۴ تا ۶۳.۲ درجه طول شرقی قرار دارد که با مساحتی معادل ۱۶۴۸۱۹۵ کیلومتر مربع حدود ۰.۳۲ درصد خشکیهای سطح زمین را در بر گرفته است. میانگین ارتفاع کشور از سطح دریا حدود ۱۲۵۰ متر و ارتفاع میانگین فلات ایران نزدیک به ۹۰۰ متر است [۱۳].

داده‌ها

برای انجام این تحقیق ۱۹ متغیر زیست‌اقلیمی از پایگاه داده (WorldClim) با تفکیک زمانی ۲.۵ دقیقه و میانگین طول دوره ۳۰ ساله از سال ۱۹۷۰ تا ۲۰۰۰ تهیه شده است.

- 1 Lumpy Skin
- 2 Arthropod Blood Feeding
- 3 World Organization For Animal Health
- 4 Geosimulation
- 5 Ecological Niche Modelling

- 6 Species Distribution Modeling
- 7 Maximum Entropy Algorithm
- 8 Presence-Only

رگرسیون درختی boosted^{۱۱} (BRT) استفاده می‌شود. اما اگر داده‌های فقط حضور موجود باشد، از مدل MaxEnt، GARP^{۱۲}، ENFA^{۱۳}، BIOCLIM، DOMAIN استفاده می‌کنند. Maxent شباهت‌های زیادی با برخی روش‌های موجود برای مدلسازی توزیع گونه‌ها، بویژه مدل‌های خطی تعمیم‌یافته (GLMs)، مدل‌های جمعی تعمیم‌یافته (GAM) (s) و روش‌های یادگیری ماشین از جمله روش‌های Bayesian و شبکه‌های عصبی دارد [۱۶]. عملکرد پیش‌بینی مدل Maxent به طور مداوم با روش‌های دیگر در حال رقابت است. از سال ۲۰۰۴ که این مدل در دسترس قرار گرفته، به طور گسترده در مدلسازی توزیع گونه‌ها به کار می‌رود. با توجه به این که در اکثر مناطق داده‌های عدم حضور محدود هستند و اکثراً بصورت فقط-حضور ثبت شده است. بنابراین استفاده از روش‌هایی که با داده‌های فقط-حضور کار کند ضروری است. مدل حداکثر آنتروپی یکی از این روش‌ها است که با اطلاعات کم و فقط - حضور کار می‌کند. و مدلسازان را از مشکل داده‌های عدم حضور غیر معتبر رها می‌سازد [۱۷-۱۸].

عملکرد مدل حداکثر آنتروپی

مدل MaxEnt داده‌های حضور را همراه با مجموعه از پیش‌بینی کننده‌های محیطی (به عنوان مثال دما، بارش) به صورت رستر و طبقه‌ای از منطقه مطالعاتی به عنوان داده‌های ورودی دریافت نموده، از همان منطقه یک نمونه از داده‌های پیش‌زمینه (background) را که جدا از محل حضور هستند و به عنوان نقاط مشکوک شناخته می‌شوند، استخراج می‌کند. داده‌های عدم حضور به این دلیل مطلوب هستند که اجازه پیش‌بینی احتمال حضور را فراهم می‌سازند و بدون داده‌های عدم حضور، پیش‌بینی معمولاً به نرخ وقوع نسبی^{۱۴} (ROR) محدود می‌شود. در اصل MaxEnt برای تخمین تراکم حضور در سراسر منطقه مطالعاتی به کار گرفته می‌شود. در صورتی که اندازه جمعیت مشخص باشد

داده‌های زیست‌اقلیمی^۱ که در مدلسازی توزیع گونه‌ها و اکولوژی مورد استفاده قرار می‌گیرد، از مقدار ماهانه دما و بارندگی استخراج شده است و نشان دهنده روند سالانه، فصلی و آستانه‌ها است. این داده‌ها توسط مرکز بین‌المللی کشاورزی حاره‌ای^۲ (CIAT) تحت برنامه تحقیقاتی تغییر اقلیم، کشاورزی و امنیت غذایی^۳ (CCAFS) به صورت صورت رایگان در دسترس عموم قرار داده شده‌اند [۱۴]. داده‌های مورد استفاده در این تحقیق شامل داده‌های اقلیمی، پارامترهای محیطی و داده‌های وقوع بیماری لمبی اسکین گاوی است. داده‌های وقوع بیماری شامل ۱۷۷۸ مورد محل وقوع این بیماری برای ماه می ۲۰۱۴ تا جولای ۲۰۱۸ برای کل ایران توسط سازمان دامپزشکی کشور ثبت شده است.

مدلسازی آشیان اکولوژیکی

مدل‌های آشیان اکولوژی^۴ یا مدل‌های توزیع گونه‌جزء پرکاربردترین اصطلاحاتی هستند که بجای مدل‌های پیش‌بینی کننده مطلوبیت زیستگاه^۵ به کار گرفته می‌شود اما در این تحقیق از اصطلاح مدلسازی مطلوبیت زیستگاه استفاده می‌کنیم این مدل‌ها بین موارد حضور گونه (محل وقوع بیماری) در محل با شرایط محیطی ارتباط برقرار می‌کنند و در واقع آشیان و توزیع گونه را مدلسازی می‌کنند [۱۵]. در دو دهه اخیر پیشرفت زیادی در مدلسازی مطلوبیت زیستگاه صورت گرفته و روش‌های متعدد در دسترس است. تمایز عمده‌ای که بین این روش‌ها وجود دارد در نوع داده‌های وقوع می‌باشد که در مدل سازی استفاده می‌شود. در صورتی که داده‌های حضور و عدم حضور گونه در دسترس باشد از روش‌های رگرسیونی مانند مدل‌های خطی تعمیم یافته^۶ (GLM) یا مدل‌های افزایشی تعمیم یافته^۷ (GAM) و ترکیب درخت رگرسیونی^۸ مانند جنگل تصادفی^۹ (RF) یا

- 1 Bioclimatic Variables
- 2 International Center For Tropical Agriculture
- 3 Climate Change, Agriculture And Food Security
- 4 Ecology Niche
- 5 Species Distribution
- 6 Habitat Suitability
- 7 Generalized linear Models
- 8 Generalized Additive Models
- 9 Ensembles of Regression Trees
- 10 Random Forests

- 11 Boosted Regression Trees
- 12 The Genetic Algorithm for Rule-Set Prediction
- 13 Environmental-Niche Factor Analysis
- 14 Relative Occurrence Rate

همراه متغیرهای ارتفاع و پوشش گیاهی برای مدلسازی انتخاب شدند (جدول ۱). برای محاسبه پیش بینی وقوع این بیماری از نرم افزار MaxEnt نسخه ۳,۴ استفاده گردید. برای ارزیابی دقت مدل نیز از محاسبه مساحت زیر منحنی ROC^۳ استفاده شد.

نتایج

شکل ۱ نقشه احتمال رخداد بیماری لمپی اسکین حاصل از اجرای مدل MaxEnt را نشان می‌دهد. بر اساس این نقشه مناطقی که شرایط زیستگاه برای وقوع این بیماری مطلوب است (احتمال بالای رخداد بیماری) با رنگ قرمز مشخص شده است. غرب و شمال غربی ایران، جنوب دریای مازندران و شامل شرقی ایران بیشترین احتمال وقوع بیماری لمپی اسکین را دارا می‌باشند.

مدل MaxEnt به پیش بینی کننده هایی که در برازش مدل نقش دارند مقادیری اختصاص می‌دهد. این مقادیر به صورت درصد مشارکت برای هر پیش بینی کننده در جدول (۲) نشان داده شده است. همان طور که مشاهده می‌شود متغیر پیش بینی کننده بارش سالانه و شاخص پوشش گیاهی در پایان فرایند کالیبراسیون مدل درصد مشارکت بیشتری نسبت به سایر متغیرهای پیش بینی کننده داشته‌اند.

جدول ۲- درصد تاثیر و مشارکت پیش بینی کننده های محیطی

مستقل در مدل MaxEnt

متغیر	درصد مشارکت
bio_12	43.3
NDVI	36.7
DEM	7.6
bio_15	5.2
bio_4	3.6
bio_2	1.7
bio_14	0.7
bio_8	0.6
bio_1	0.5

این مدل نرخ وقوع واقعی را پیش بینی می‌کند که به عنوان تعداد موارد مورد انتظار در آن سلول تعریف می‌شود. اما معمولاً اندازه جمعیت مشخص نیست، در این صورت تنها یک مقایسه نسبی میان نرخ‌ها معنی دار است و منجر به محاسبه نرخ وقوع نسبی می‌شود. در این صورت احتمال حضور با استفاده از تبدیل لجستیک نرخ وقوع نسبی محاسبه می‌شود (رابطه ۱). در این رابطه نرخ وقوع نسبی تابعی از پیش بینی کننده های محیطی در نقاط حضور است.

$$p * (z(x_i)) = \exp(z(x_i)\lambda) / \sum_i \exp(z(x_i)\lambda) \quad (1)$$

که در رابطه فوق

(2)

$$z(x_i)\lambda = z_1(x_i) * \lambda_1 + z_2(x_i) * \lambda_2 + \dots + z_r(x_i) * \lambda_r$$

در رابطه (۲)، Z بردار متغیرهای پیش بینی کننده i مانند (دما، بارش، ارتفاع...) در محل و λ بردار ضرایب رگرسیون است. معادله (۱) یک کلاس کلی از مدل‌هایی است که برای پیش بینی نرخ وقوع نسبی طراحی شده و به درک مدل Maxent از دو دیدگاه آماری و یادگیری ماشین کمک می‌کند [۱۷].

روش تحقیق

به منظور انجام تحقیق حاضر از دو دسته متغیرهای مستقل زیست اقلیمی^۱ (۱۹ متغیر) و متغیرهای شاخص پوشش گیاهی و مدل رقومی ارتفاع و همنچنین متغیر وابسته‌ی وقوع بیماری لمپی اسکین، استفاده شد. برای انتخاب متغیرهای موثر بر توزیع مکانی بیماری لمپی اسکین از ضریب همبستگی اسپیرمن بین متغیرها استفاده شد. روابط بین متغیرهای پیش‌بینی کننده بررسی و متغیرهایی که دارای ضریب همبستگی $-0.8 < r < 0.8$ و بیشترین درصد اهمیت نسبی بر اساس آزمون جک نایف^۲ بودند، به عنوان پیش بینی کننده‌های محیطی برای مدلسازی وقوع این بیماری انتخاب شدند. بر این اساس از ۱۹ متغیر زیست اقلیمی موجود، ۷ متغیر (۴ متغیر دمایی و ۳ متغیر بارشی) غیرهمبسته و با بیشترین درصد اهمیت نسبی در آزمون به

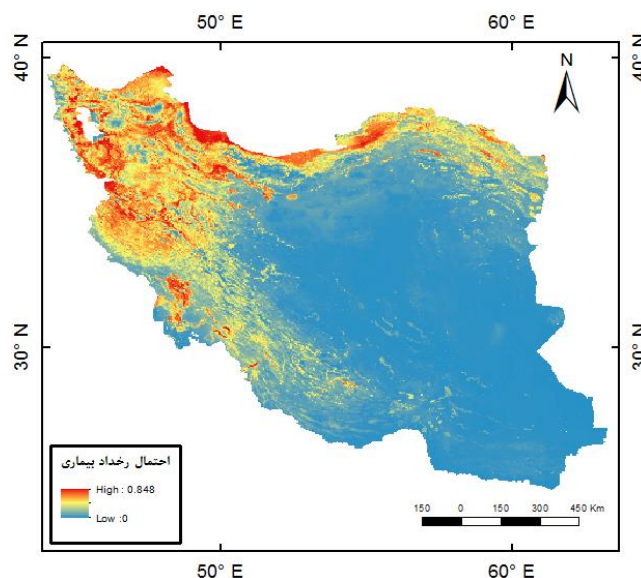
1 Bioclimatic

2 Jackknife

3 Receiver Operating Characteristic

جدول ۱- متغیرهای پیش‌بینی‌کننده منتخب

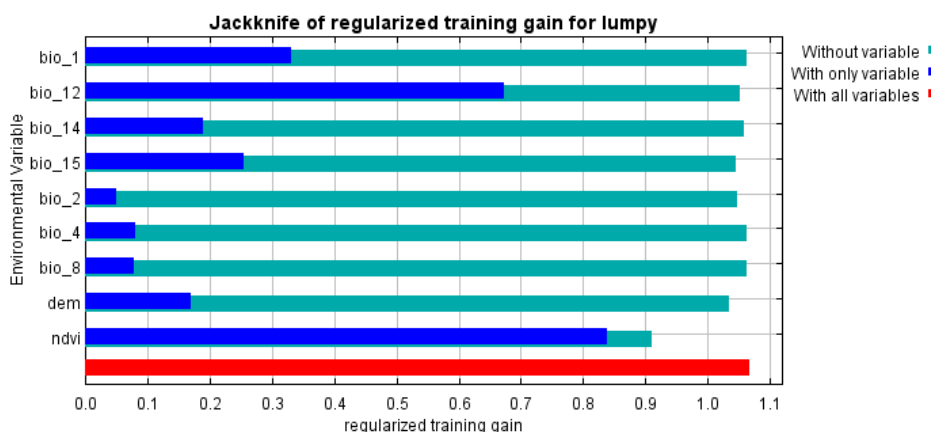
ردیف	متغیر	نام اختصاری
۱	دمای متوسط سالانه	Bio_1
۲	دامنه میانگین دمای روزانه	Bio_2
۳	رژیم دما	Bio_3
۴	میانگین دمای مرطوب‌ترین فصل	Bio_8
۵	بارش سالانه	Bio_12
۶	بارش خشک‌ترین ماه	Bio_14
۷	رژیم بارش	Bio_15
۸	ارتفاع	DEM
۹	شاخص گیاهی تفصیلی نرمال شده	NDVI



شکل ۱- نقشه احتمال رخداد بیماری لمپی اسکین در ایران

متغیری که بیشترین تاثیر در بهبود عملکرد مدل را دارد به عنوان مهمترین متغیر تاثیر گذار انتخاب می‌شود. باتوجه به شکل (۲) مانند حالت پیشین شاخص‌های پوشش گیاهی و بارش سالانه موثرترین پیش‌بینی‌کننده‌ها برای توزیع داده‌های وقوع بیماری لمپی اسکین گاوی در ایران به شمار می‌روند.

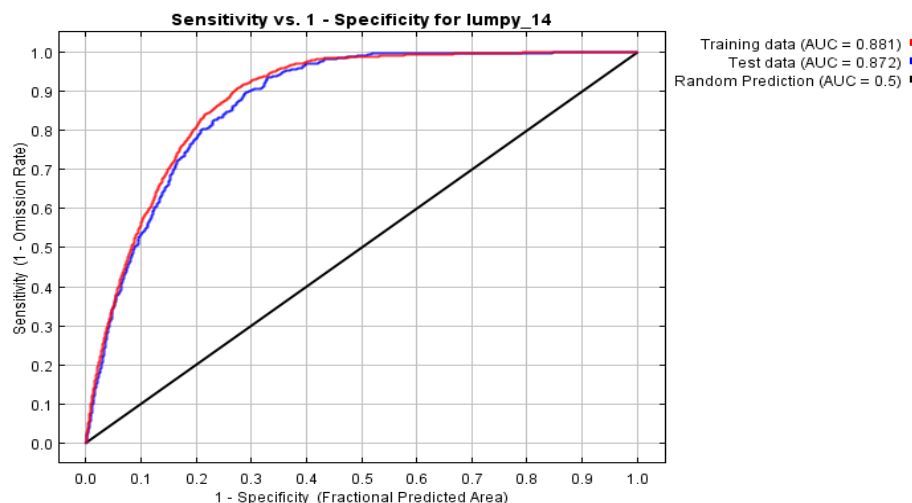
مدل Maxent برای تعیین اهمیت هر متغیر پیش‌بینی‌کننده از آزمون جک‌نایف نیز استفاده می‌کند. در این آزمون هر بار یک متغیر پیش‌بینی‌کننده از مدل خارج می‌شود و مدل با استفاده از سایر پیش‌بینی‌کننده‌های باقی‌مانده اجرا می‌شود و متغیر پیش‌بینی‌کننده‌ای که بیشتر از سایر متغیرها کارایی مدل را پایین می‌آورد، حذف می‌شود و در نهایت



شکل ۲- آزمون جک نایف برای بررسی اهمیت پیش بینی کننده های مستقل در مدل MaxEnt

باشد. این نتایج نشان می دهد علاوه بر اینکه دقت مدل ها در مقایسه با حالت تصادفی که در آن $AUC=0.5$ می باشد به میزان قابل توجهی بالاتر است؛ با توجه به اختلاف اندک مقادیر AUC در داده های آموزشی و تست، مدل دارای مشکل بیش برزش نیز نمی باشد.

در این تحقیق از مساحت زیر منحنی ROC - که اصطلاحاً AUC نامیده می شود- برای ارزیابی دقت پیش بینی مدل استفاده شده است. در شکل (۳) مقدار AUC برای داده های آموزشی و داده های آزمون برآورد شده است. همانطور که مشاهده می شود این مقدار در بهترین اجرا برای داده های آموزشی برابر ۰.۸۸ و برای داده های آزمون ۰.۸۷ می



شکل ۳- منحنی ROC و حالت های مختلف پیش بینی کننده در آن

بکارگیری مدل MaxEnt بر مبنای داده‌های " فقط-حضور " و متغیرهای زیست اقلیمی، امکان مدلسازی احتمال رخداد بیماری لمپی اسکین را در سطح کشور فراهم می‌سازد با توجه به این که این مدل مناطقی را شناسایی می کند که تا

نتیجه گیری

این مطالعه یک رویکرد جدید را برای مدلسازی مکانی بیماری لمپی اسکین در کشور ایران فراهم می کند..

9. Stevens, K., Pfeiffer, D., 2011. *Spatial modelling of disease using data- and knowledge-driven approaches. Spatial and Spatio-temporal Epidemiology* 2, 125–133.

10. Jane Elith, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17:43-57.

11. Phillips SJ, Andreson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Modell* (2006) 190(3–4):231–59.

12. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* (2006) 29(2):129–51

۱۳. علایی طالقانی، محمود، 5985 و 5983 ، ژئومورفولوژی ایران، چاپ اول و چاپ هفتم، نشر قومس

14. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *JOC Int J Climatol* (2005) 25(15):1965–78.

15. Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.

16. Phillips SJ, Andreson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Modell* (2006) 190(3–4):231–59.

17. Corey Merow, Matthew J. Smith, and John A. Silander, Jr. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36: 1058–1069.

18. Busby, J.R., 1986. A biogeographical analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Aust. J. Ecol.* 11, 1–7.

بحال گزارشی از این بیماری در آنها ثبت نشده است؛ در نتیجه، مدلسازی مکانی احتمال رخداد بیماری‌ها گامی حیاتی در جهت اتخاذ استراتژی‌های مؤثر برای ارتقاء بهداشت عمومی به شمار می‌رود.

منابع

1. Abera Z, Degefu H, Gari G, Kidane M. Sero-prevalence of lumpy skin disease in selected districts of West Wollega zone, Ethiopia. *BMC Vet Res* (2015) 11:135.

2. MacDonald RAS. *Pseudo-Urticaria of Cattle*. Government of Northern Rhodesia: Department of Animal Health (1931). p. 20–1.

3. Davies FG. Observations on the epidemiology of lumpy skin disease in Kenya. *J Hyg (Lond)* (1982) 88(1):95–102.

4. Burdin ML, Prydie J. Lumpy skin disease of cattle in Kenya. *Nature* (1959) 183(4666):949–50

5. Gari G, Bonnet P, Roger F, Waret-Szkuta A. Epidemiological aspects and financial impact of lumpy skin disease in Ethiopia. *Prev Vet Med* (2011) 102(4):274–83

6. Sameea Youseei, P., Mardani, K., Dalir-Naghadeh, B. & Jalilzadeh-Amin, G. 2017. Epidemiological Study of Lumpy skin disease Outbreaks in North-western Iran. *Transboundary and emerging diseases*, 64, 1782-1789.

7. Cingoli, G. & Rinaldi, L. 2007. Geographical Information Systems in buffalo health applications. *Italian Journal of Animal Science*, 6, 217-222.

8. Chen, D., Moulin, B. & WU, J. 2014. *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, Wiley