

مدل‌سازی و پیش‌بینی فضایی بروسولوز انسانی در سبزوار با استفاده از یادگیری ماشین:

مقایسه دو روش PCA-SVM و PCA-MLP

ایوب رستگار^۱، علی اوغازیان^۲، شیمیا امیری پارسا^۳، علی پوریوسف^۴، محمدشفیق مجددی^۵

مقاله پژوهشی

چکیده

مقدمه: بیماری بروسولوز (تب ملت)، یک بیماری باکتریایی مشترک بین انسان و دام است که در مناطق مختلفی از ایران از جمله شهرستان سبزوار شیوع دارد. هدف از این مطالعه، مدل‌سازی و پیش‌بینی فضایی بروز بروسولوز انسانی با استفاده از سیستم اطلاعات جغرافیایی (GIS) و الگوریتم‌های یادگیری ماشین شامل شبکه‌ی عصبی پرسپترون چندلایه همراه با تحلیل مؤلفه‌های اصلی (PCA-MLP) و ماشین بردار پشتیبان همراه با تحلیل مؤلفه‌های اصلی (PCA-SVM) در این شهرستان بود.

روش‌ها: داده‌های مربوط به بروز بروسولوز، اطلاعات جمعیتی، اقلیمی، پوشش گیاهی و توپوگرافی بین سال‌های ۱۳۹۰ تا ۱۴۰۱ جمع‌آوری شد. داده‌ها با استفاده از GIS پردازش و سپس با دو الگوریتم یادگیری ماشین MLP و SVM همراه با تحلیل مؤلفه‌های اصلی (PCA) جهت مدل‌سازی و پیش‌بینی توزیع فضایی بیماری به کار گرفته شدند.

یافته‌ها: در این مطالعه، تحلیل فصلی نشان داد که در سال‌های ۱۳۹۱، ۱۳۹۳-۱۳۹۶، ۱۳۹۹ و ۱۴۰۱ اوج بروز بروسولوز در فصل بهار رخ داده، در حالی که در سال‌های ۱۳۹۸ و ۱۴۰۰ این اوج به تابستان منتقل شده و در برخی سال‌ها (۱۳۹۰، ۱۳۹۲، ۱۳۹۷) توزیع فصلی بیماری روندی ناپایدار داشته است. همچنین رطوبت هوا و ارتفاع از سطح دریا به‌عنوان مهم‌ترین پیش‌بینی‌کننده‌های بروز شناخته شدند. مدل PCA MLP مناطقی مانند ششتمد و دستوران را با بیشترین تعداد موارد انسانی پیش‌بینی کرد، در حالی که مدل PCA SVM علاوه بر این‌ها، برخی مناطق دیگر مورد مطالعه را نیز در گروه مناطق پرخطر قرار داد.

نتیجه‌گیری: الگوریتم PCA-SVM عملکرد دقیق‌تری در مدل‌سازی و پیش‌بینی بروز بیماری نسبت به PCA-MLP داشت و مناطق با ریسک بالا مانند ششتمد و دستوران را دقیق‌تر شناسایی کرده بود. یافته‌های این مطالعه می‌تواند به مدیران سلامت در تخصیص منابع و طراحی مداخلات پیشگیرانه در مناطق پرخطر کمک کند.

واژگان کلیدی: بروسولوز؛ سیستم‌های اطلاعات جغرافیایی؛ الگوریتم‌های یادگیری ماشین؛ مدل‌های شبکه عصبی؛ تحلیل مؤلفه‌های اصلی؛ پیش‌بینی

ارجاع: رستگار ایوب، اوغازیان علی، امیری پارسا شیمیا، پوریوسف علی، مجددی محمدشفیق. مدل‌سازی و پیش‌بینی فضایی بروسولوز انسانی در سبزوار با

استفاده از یادگیری ماشین: مقایسه‌ی دو روش PCA-SVM و PCA-MLP. مجله دانشکده پزشکی اصفهان ۱۴۰۵؛ ۴۴ (۸۵۴): ۳۴۸-۳۵۷.

مقدمه

عضلانی و مفصلی همراه است که تشخیص بالینی آن را با چالش مواجه می‌سازد. بروسولوز در صورت عدم تشخیص و درمان به موقع می‌تواند منجر به عوارض مزمن و حتی ناتوان‌کننده در افراد شود (۱، ۲).

در ایران، به‌ویژه در نواحی روستایی و عشایری، به‌دلیل فعالیت‌های دامداری گسترده، نبود کنترل کافی بر فرآورده‌های دامی و عادات غذایی سنتی، بروسولوز به‌عنوان یک معضل بهداشت عمومی شناخته می‌شود

بروسولوز (تب ملت)، یکی از مهم‌ترین بیماری‌های زئونوزها منشأ باکتریایی در جهان است که سالانه صدها هزار نفر را درگیر می‌سازد. این بیماری عمدتاً از طریق تماس مستقیم با دام آلوده یا مصرف محصولات لبنی غیرپاستوریزه به انسان منتقل می‌شود و با طیف گسترده‌ای از علائم غیراختصاصی مانند تب مواج، تعریق شبانه، درد

۱- دانشیار، گروه مهندسی بهداشت محیط، دانشگاه علوم پزشکی خراسان شمالی، بجنورد، ایران

۲- دانشجوی دکتری بهداشت محیط، گروه مهندسی بهداشت محیط، دانشکده‌ی بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۳- دانشجوی پزشکی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی سبزوار، سبزوار، ایران

۴- استادیار انگل‌شناسی، مرکز تحقیقات لیشمانیوز، دانشگاه علوم پزشکی سبزوار، سبزوار، ایران

۵- استاد ایمنولوژی، مرکز تحقیقات لیشمانیوز، دانشگاه علوم پزشکی سبزوار، سبزوار، ایران

نویسنده‌ی مسؤول: محمدشفیق مجددی؛ استاد ایمنولوژی، مرکز تحقیقات لیشمانیوز، دانشگاه علوم پزشکی سبزوار، سبزوار، ایران

روش‌ها

منطقه مورد مطالعه:

شهرستان سبزوار، واقع در استان خراسان رضوی، با مساحت ۲۰،۴۸۴ کیلومتر مربع، طول و عرض جغرافیایی ۳۶°۲۱'۵۱" و ۵۷°۶۶'۷۸" و ارتفاع متوسط ۹۷۷ متر از سطح دریا، در غرب این استان قرار دارد. این شهرستان با جمعیتی بالغ بر ۲۹۸،۵۸۸ نفر، شامل چهار بخش مرکزی، جوین، جغتای و خوشاب است (شکل ۱). از نظر اقلیمی، سبزوار دارای اقلیم نیمه‌خشک با تابستان‌های گرم و خشک و زمستان‌های سرد و مرطوب است میانگین دمای سالانه این شهرستان حدود ۱۶ درجه‌ی سانتی‌گراد بوده و میانگین بارندگی سالانه آن حدود ۱۸۷ میلی‌متر است.

پس از دریافت کد اخلاق از کمیته اخلاق دانشگاه علوم پزشکی سبزوار (IR.MEDSAB.REC.1401.061)، داده‌های مربوط به بروز بروسلوز انسانی و عوامل مؤثر بر آن، شامل اطلاعات سرشماری جمعیت، اقلیم، پوشش گیاهی و ویژگی‌های توپوگرافی، از منابع معتبر جمع‌آوری شد. برای پردازش و تحلیل این داده‌ها، از نرم‌افزارهای Topo 2016 و ArcGIS 10.2 استفاده گردید. همچنین، ابزار Raster to ArcGIS در ArcGIS برای درون‌یابی داده‌های توپوگرافی مانند دما و بارش به‌کار گرفته شد.

داده‌های بیماری و جمعیت

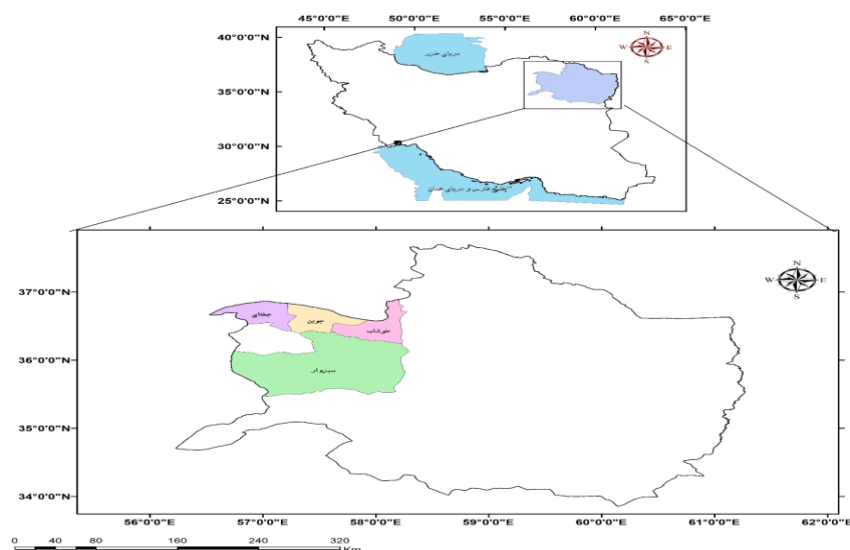
اطلاعات مربوط به جمعیت محل مورد مطالعه و میزان بروز ماهانه بروسلوز بین سال‌های ۱۳۹۰ تا ۱۴۰۱ به ترتیب از مرکز آمار ایران و معاونت بهداشتی دانشگاه علوم پزشکی سبزوار اخذ شد.

داده‌های آب و هوا

داده‌های اقلیمی شامل تخیخ ماهانه کل، حداکثر سرعت باد، تعداد روزهای آفتابی، میانگین دما، تعداد روزهای یخبندان و رطوبت

(۳). شهرستان سبزوار به‌دلیل موقعیت جغرافیایی خاص، اقلیم نیمه‌خشک، وجود دامداری‌های سنتی، و مصرف فراورده‌های لبنی محلی از مناطق یا شیوع بالای بیماری به‌شمار می‌آید. با توجه به پیچیدگی‌های الگوی انتقال بروسلوز و اثرپذیری آن از عوامل متعدد محیطی، جمعیتی و اقلیمی، استفاده از ابزارهای نوین مانند سامانه اطلاعات جغرافیایی (GIS) در کنار مدل‌های یادگیری ماشینی، امکان تحلیل فضایی-زمانی دقیق‌تر این بیماری را فراهم کرده است (۴).

GIS با تجمیع و تحلیل داده‌های مکانی می‌تواند خوشه‌های پرخطر را شناسایی و پیش‌بینی کند و به برنامه‌ریزی هدفمند در کنترل بیماری کمک کند. در سال‌های اخیر، الگوریتم‌های یادگیری ماشینی به‌ویژه مدل‌های ترکیبی مانند شبکه عصبی پرسپترون چندلایه همراه با تحلیل مؤلفه‌های اصلی (PCA-MLP) و ماشین بردار پشتیبان همراه با تحلیل مؤلفه‌های اصلی (PCA-SVM) به دلیل قدرت بالا در مدیریت داده‌های چندبعدی و غیرخطی، در اپیدمیولوژی بیماری‌ها مورد توجه قرار گرفته‌اند (۵-۸). ترکیب PCA با مدل‌های یادگیری ماشینی، علاوه بر کاهش ابعاد داده‌ها و حذف هم‌خطی، باعث بهبود دقت پیش‌بینی می‌شود. در مطالعات مختلف، این ترکیب توانسته عملکرد مدل‌ها را در تحلیل توزیع بیماری‌هایی مانند لیشمانیوز، توبرکلوزیس، و لپتوسپیروز ارتقاء دهد (۵، ۷، ۹). با توجه به خلأ موجود در مطالعات جامع مکانی بروسلوز انسانی در منطقه سبزوار، این مطالعه با هدف مدلسازی و پیش‌بینی فضایی این بیماری با استفاده از GIS و الگوریتم‌های یادگیری ماشینی PCA-MLP و PCA-SVM انجام شده است. ارائه الگویی دقیق برای شناسایی مناطق پرخطر، تسهیل تصمیم‌سازی برای مداخلات بهداشتی و ارتقاء سطح آمادگی نظام سلامت در مواجهه با بیماری‌های زئونوز، از کاربرد نتایج این مطالعه می‌باشد.



شکل ۱. موقعیت جغرافیایی شهرستان سبزوار

Backpropagation) و گرادینان نزولی آموزش می‌بیند تا وزن‌ها را به‌روزرسانی کند و خطای مدل را به حداقل برساند.

یافته‌ها

الگو و عوامل مؤثر بر پراکندگی بیماری بروسلوز در سال‌های گوناگون
بر اساس تحلیل شاخص موران I، موارد بروسلوز در سال‌های ۱۳۹۰، ۱۳۹۱، ۱۳۹۷-۱۳۹۳، ۱۳۹۹ و ۱۴۰۱ به صورت خوشه‌ای و در سال‌های ۱۳۹۲، ۱۳۹۸ و ۱۴۰۰ به صورت تصادفی توزیع شده بودند. آمارهای توصیفی تکمیلی و داده‌های NDVI در جدول ۱ به تفصیل آمده است. رابطه‌ی بین عوامل بررسی‌شده و ابتلا به بروسلوز متفاوت بود؛ به این صورت که بر اساس آزمون Pearson شیب، NDVI، رطوبت، تعداد روزهای یخبندان و ارتفاع با بروز بروسلوز همبستگی منفی داشتند، در حالی که تعداد روزهای آفتابی، دما، سرعت باد و تبخیر با بروز بروسلوز همبستگی مثبت داشتند.

جدول ۱. بررسی یافته‌های توصیفی

داده‌های توصیفی	رنج	میانگین	کمترین	بیشترین
شیب (m/s)	۰/۲۶	۰/۱۹۰۰	۰/۰۹	۰/۳۵
NDVI	۰/۰۷	۰/۳۳۵۰	۰/۳۰	۰/۳۷
رطوبت	۰/۰۹	۰/۲۶۰۰	۰/۲۱	۰/۳۰
سرعت باد (Km/h)	۱۱/۰۰	۲۸/۵۰۰۰	۲۵/۰۰	۳۶/۰۰
دما (c0)	۵/۰۰	۲۲/۰۰۰۰	۱۹/۰۰	۲۴/۰۰
ارتفاع (m)	۱۷۱۳/۰۰	۱۵۵۶/۷۵۰۰	۹۷۷/۰۰	۲۶۹۰/۰۰
روزهای یخبندان	۱۲/۰۰	۳۱/۲۵۰۰	۲۶/۰۰	۳۸/۰۰
روزهای آفتابی	۳۴/۰۰	۹۱/۵۰۰۰	۷۹/۰۰	۱۱۳/۰۰
تبخیر (kJ/mol)	۷۵۰/۰۰	۲۱۲۵/۰۰۰۰	۱۷۵۰/۰۰	۲۵۰۰/۰۰

دقت مدل‌های مختلف در پیش‌بینی بروز بروسلوز انسانی

بر اساس نتایج AUC، مدل PCA-SVM عملکرد بهتری نسبت به PCA-MLP داشت. افزون بر آن، با توجه به مقادیر حاصل از MAE، R² و RMSE در جدول ۲، مشخص شد که PCA-SVM از نظر قابلیت اطمینان در پیش‌بینی موارد بیماری، مدل برتری است.

جدول ۲. ارزیابی تکنیک‌های PCA-MLP و PCA-SVM برای پیش‌بینی موارد بروسلوز با استفاده از داده‌های تست در مرحله‌ی توسعه مدل

(۱۳۹۰-۱۴۰۱)

الگوریتم	R2	MAE	RMSE
PCA-MLP	۰/۸۱۳۷	۰/۳۱	۳۶/۷۶
PCA-SVM	۰/۸۸۲۹	۰/۲۵	۲۸/۴۵

بین سال‌های ۱۳۹۰ تا ۱۴۰۱ از سازمان هواشناسی کشور (<https://data.irimo.ir/>) جمع‌آوری شد.

داده‌های توپوگرافی و پوشش گیاهی

برای تحلیل ویژگی‌های توپوگرافی، نقشه‌های ارتفاعی از سازمان زمین‌شناسی ایالات متحده (USGS) و تصاویر ماهواره‌ای ASTER با وضوح ۹۰ متر استفاده شد. این داده‌ها به منظور بررسی تأثیرات ارتفاع بر بروز بیماری در مناطق مختلف شهرستان سبزوار به کار گرفته شدند. شاخص نرمال‌شده تفاوت پوشش گیاهی (NDVI) از تصاویر ماهواره‌ای MODIS در بازه زمانی ۱۳۹۰ تا ۱۴۰۱ استخراج گردید. این شاخص به منظور بررسی وضعیت پوشش گیاهی و ارتباط آن با بروز بیماری در مناطق مختلف شهرستان سبزوار مورد استفاده قرار گرفت.

پردازش داده‌ها و تحلیل فضایی

داده‌های استخراج‌شده شامل NDVI، توپوگرافی و عوامل اقلیمی به صورت شطرنجی درآمده و مقادیر میانگین پیکسل‌ها در هر منطقه روستایی با استفاده از تحلیل فضایی منطقه‌ای محاسبه شد. این فرایند به منظور آماده‌سازی داده‌ها برای مدل‌سازی و پیش‌بینی بروز بیماری انجام شد.

آزمون‌های آماری

برای ارزیابی مناسب بودن داده‌ها جهت تحلیل مؤلفه‌های اصل (PCA)، آزمون‌های بارتلت و KMO انجام شد. آزمون بارتلت برای بررسی کافی بودن نمونه‌ها در تحلیل عاملی اکتشافی و آزمون KMO به‌عنوان معیاری برای کافی بودن اندازه نمونه در نظر گرفته شد.

تحلیل مؤلفه‌های اصلی (PCA)

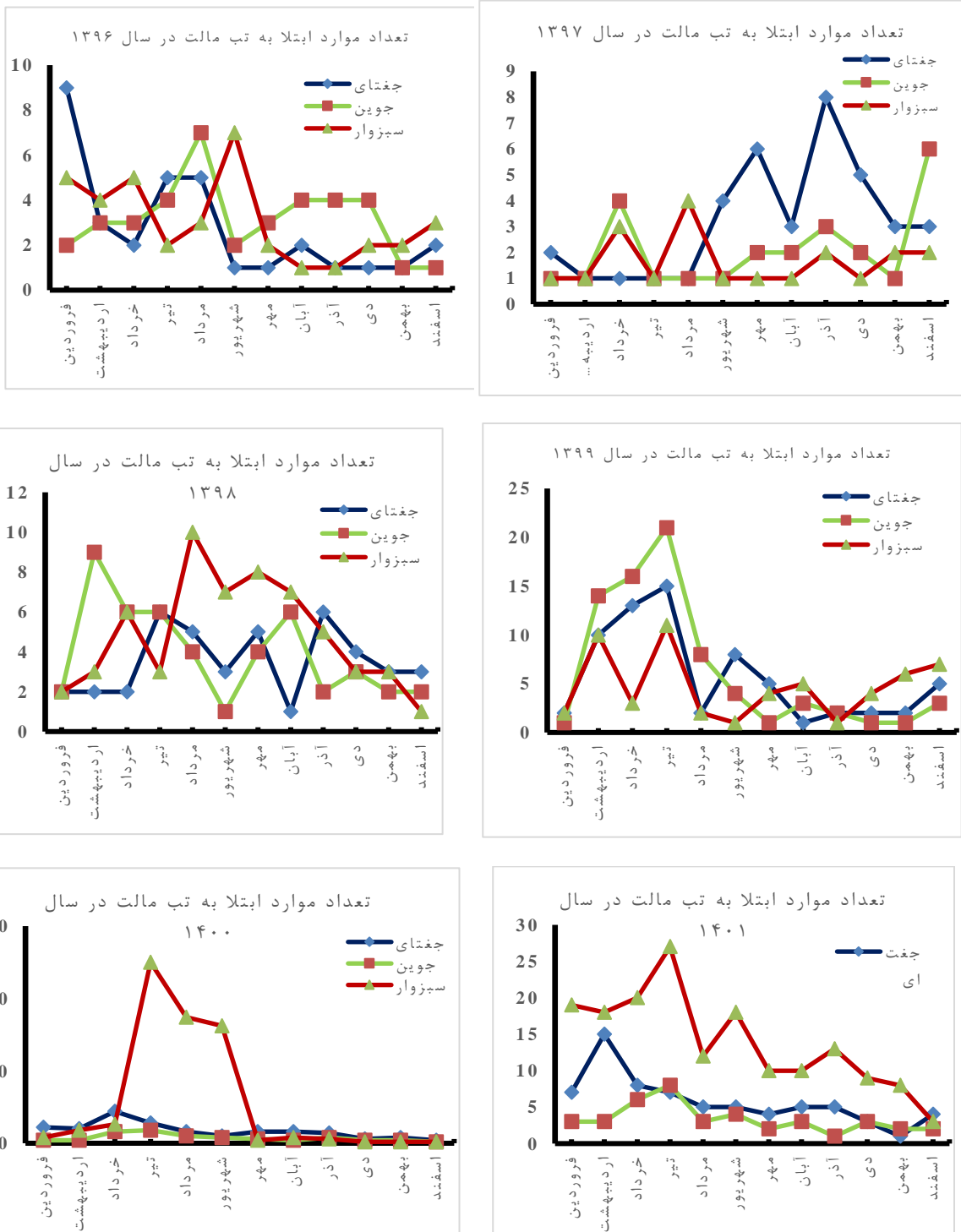
روش PCA به منظور کاهش ابعاد داده‌ها و استخراج مؤلفه‌های اصلی استفاده شد. این روش با به حداکثر رساندن همبستگی بین پارامترهای اصلی، متغیرهای جدیدی به نام PC را تشکیل می‌دهد و پیچیدگی داده‌ها را کاهش می‌دهد.

ماشین بردار پشتیبان (SVM)

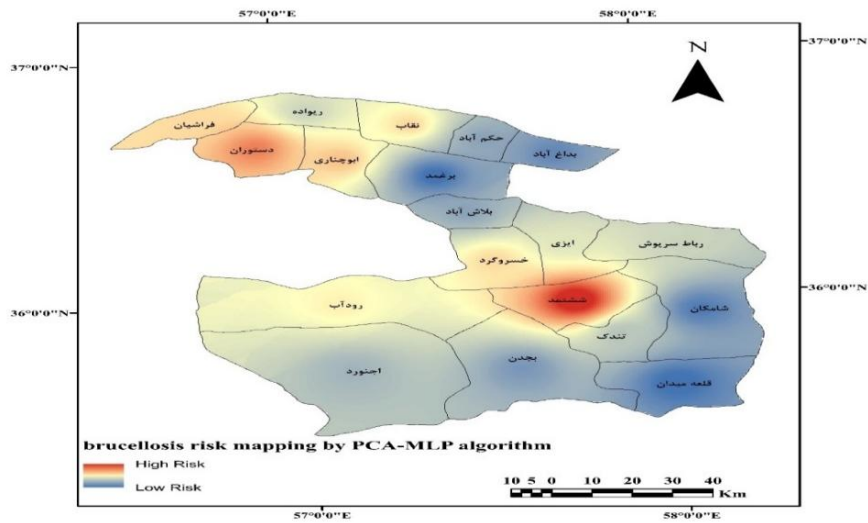
یک روش یادگیری نظارت‌شده برای مسائل طبقه‌بندی و رگرسیون است که با یافتن یک ابرصفحه (Hyperplane) بهینه، داده‌ها را طوری جدا می‌کند که حداکثر فاصله (Margin) ممکن بین نزدیک‌ترین نمونه‌های هر کلاس حفظ شود. این ویژگی باعث افزایش توانایی مدل در تعمیم‌پذیری و مقابله با داده‌های نویزی می‌شود.

پرسترون چندلایه (MLP)

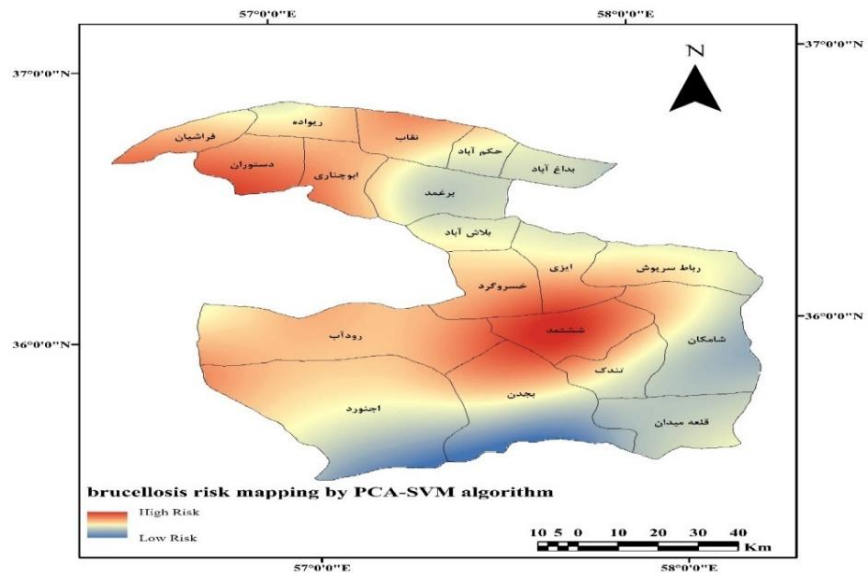
نوعی شبکه‌ی عصبی پیش‌خور است که شامل یک یا چند لایه پنهان بین ورودی و خروجی است و توسط نورون‌هایی با توابع فعال‌سازی غیرخطی، توانایی مدل‌سازی روابط پیچیده و غیرخطی بین داده‌ها را دارد. این شبکه با استفاده از روش پس‌انتشار خطا



شکل ۳. روند تغییرات بیماری بروسلوز انسانی در فصل‌های مختلف از سال ۱۳۹۶ تا ۱۴۰۱ در مناطق مورد مطالعه



شکل ۴. نقشه‌ی پیش‌بینی ریسک خطر بروسلوزیس با مدل PCA-MLP



شکل ۵. نقشه‌ی پیش‌بینی ریسک خطر بروسلوزیس با مدل PCA-SVM

ایزی را نیز به‌عنوان مناطقی با خطر بالا برای بروسلوز انسانی پیش‌بینی نمود (شکل ۵).

بحث

در این پژوهش، با استفاده از دو الگوریتم یادگیری ماشین-PCA-MLP و PCA-SVM، اقدام به مدلسازی و پیش‌بینی فضایی بروز بیماری بروسلوز انسانی در شهرستان سبزوار نمودیم. دارا بودن ارتفاعات مختلف از سطح دریا، باعث ایجاد اقلیم‌های متنوع در منطقه مورد مطالعه شده‌اند، لذا پیش‌بینی این بیماری که به شدت تحت تأثیر

مدلسازی و پیش‌بینی میزان بروز بروسلوز انسانی در سطح

شهرستان سبزوار، با استفاده از دو روش PCA-SVM و PCA-MLP

با استفاده از مدل PCA-MLP، بیشترین میزان ابتلای پیش‌بینی شده به بروسلوز انسانی مربوط به بخش ششتمد و پس از آن دستوران بود (شکل ۴). با مدل PCA-SVM نیز کانون‌های اصلی بیماری همچنان ششتمد و دستوران تشخیص داده شدند؛ اما برخلاف مدل PCA-MLP، این مدل مناطقی مانند ابوچناری، نقاب، فراشیان، روداب، خسروگرد، بخش‌های شمالی بجدن و تندک و بخش جنوبی

شبکه‌ی عصبی مصنوعی جدید (ANN) با استفاده از تعداد سالانه موارد بروسولوز انسانی در چین، توسعه داده شد. داده‌های سال‌های ۲۰۰۶ تا ۲۰۰۸ برای آموزش، اعتبارسنجی و آزمایش مدل استفاده شد، در حالی که داده‌های سال‌های ۲۰۰۹-۲۰۱۰ برای ارزیابی عملکرد مدل استفاده شد. نتایج نشان داد که شاخص گیاهی تقویت شده به عنوان مهم‌ترین پیش‌بینی‌کننده بروز بروسولوز انسانی شناسایی شد و پس از آن دمای سطح زمین و سایر متغیرهای مرتبط با دما و بارش قرار گرفتند (۴).

باتوجه به نتایج بدست آمده از مطالعه‌ی ما، شاخص رطوبت به عنوان مهم‌ترین پیش‌بینی‌کننده بروز بروسولوز انسانی شناسایی شد و در مرتبه‌ی بعدی شاخص ارتفاع قرار گرفت. در مورد ریسک خطر بروسولوز انسانی و شناسایی مناطق با خطر بالا یعنی (ششتمد و دستوران) هر دو مدل PCA-SVM و PCA-MLP پیش‌بینی نزدیک بهم داشتند. اما دقت مدل PCA-SVM نسبت به مدل دیگر بالاتر بود. این مطالعه همسو با مطالعه‌ی Wang و همکاران بود و نشان داد که با استفاده از مدل‌های PCA-SVM و PCA-MLP و متغیرهای محیطی به‌دست‌آمده از داده‌های ماهواره‌ای، می‌توان بروز بروسولوز انسانی را با درجه‌ای از دقت معقول پیش‌بینی کرد.

Tabasi و همکاران در سال ۲۰۲۰، مطالعه‌ای با عنوان مدلسازی فضایی لیسمانیوز جلدی با توجه به عوامل محیطی بالقوه با استفاده از روش‌های عصبی فازی استنتاج، تطبیقی انجام دادند. نتایج این مطالعه بیانگر این بود که مدل PCA-ANFIS با مقادیر $RMSE = 0.791$, $R2 = 0.705$, $MAE = 0.827$ و $MAE = 0.904$ در مقایسه با مدل ANFIS با مقادیر $RMSE = 0.791$, $R2 = 0.705$, $MAE = 0.827$ و $MAE = 0.904$ عملکرد بهتری در پیش‌بینی داشت. نقشه‌های واقعی و پیش‌بینی شده موارد ZCL توسط هر دو مدل نشان می‌دهد که مناطق پرخطر بیماری در شمال شرق، شمال و برخی از دهستان‌های مرکزی استان گلستان قرار دارند. تجزیه و تحلیل حساسیت مدل ANFIS نشان داد که جمعیت، پوشش گیاهی، میانگین سرعت باد، ارتفاع و میانگین دمای خاک به ترتیب مهم‌ترین عوامل در پیش‌بینی موارد لیسمانیوز هستند. همچنین نتایج تأیید کرد که شیوع بیماری عمدتاً در مناطق بیابانی، مناطق کم ارتفاع و مراکز جمعیتی کنار رودخانه‌ها رخ می‌دهد (۷).

در مطالعه‌ی ما که به مقایسه‌ی دو روش PCA-SVM و PCA-MLP در پیش‌بینی موارد ابتلا به بروسولوز انسانی در شهرستان سبزوار پرداختیم، مدل PCA-SVM در مقایسه با مدل دیگر با مقادیر $RMSE = 28/45$ و $MAE = 0/25$, $R2 =$ پیش‌بینی داشت. از عوامل مؤثر شناخته شده در این مطالعه نیز می‌توان

شرایط محیطی است، با مشکلاتی مواجه شده است. افزایش موارد بروز بروسولوز در سراسر شهرستان عمدتاً به نحوه توزیع جغرافیایی و تغییرات پویا در پارامترهای مؤثر اعم از عوامل اجتماعی، اقتصادی، جمعیتی و محیطی وابسته است. به عبارت دیگر، تغییرات این مؤلفه‌ها در طول دوره‌ی مطالعه بر بروز بیماری اثرگذار بوده و می‌تواند بر تعداد مبتلایان تأثیر مستقیم داشته باشد. نتایج تحلیل‌ها نشان داد که مدل PCA-SVM نسبت به PCA-MLP عملکرد دقیق‌تری در پیش‌بینی خوشه‌های بیماری ارائه می‌دهد. این یافته با نتایج مطالعات مشابه در داخل و خارج کشور هم‌خوانی دارد و تأکید می‌کند که رویکردهای مبتنی بر یادگیری ماشینی و تحلیل مکانی در مطالعات اپیدمیولوژیک بیماری‌های عفونی نقش مؤثری دارند.

مطالعه‌ای در سال ۲۰۱۷ بر روی رویکرد شبکه‌ی عصبی مصنوعی مبتنی بر داده‌های سنجش از دور جهت پیش‌بینی شیوع بیماری بروسولوز انسانی انجام دادند. در این مطالعه از بروسولوز انسانی به عنوان یک مورد آزمایشی برای شناسایی عوامل مهم محیطی تعیین‌کننده بیماری و پیش‌بینی شیوع آن استفاده شد. یک مدل شبکه‌ی عصبی مصنوعی جدید (ANN) با استفاده از تعداد سالانه موارد بروسولوز انسانی در چین، توسعه داده شد. داده‌های سال‌های ۲۰۰۶ تا ۲۰۰۸ برای آموزش، اعتبارسنجی و آزمایش مدل استفاده شد، در حالی که داده‌های سال‌های ۲۰۰۹-۲۰۱۰ برای ارزیابی عملکرد مدل استفاده شد. نتایج نشان داد که شاخص گیاهی تقویت شده به عنوان مهم‌ترین پیش‌بینی‌کننده بروز بروسولوز انسانی شناسایی شد و پس از آن دمای سطح زمین و سایر متغیرهای مرتبط با دما و بارش قرار گرفتند (۴).

باتوجه به نتایج بدست آمده از مطالعه‌ی ما، شاخص رطوبت به عنوان مهم‌ترین پیش‌بینی‌کننده بروز بروسولوز انسانی شناسایی شد و در مرتبه‌ی بعدی شاخص ارتفاع قرار گرفت. در مورد ریسک خطر بروسولوز انسانی و شناسایی مناطق با خطر بالا یعنی (ششتمد و دستوران) هر دو مدل PCA-SVM و PCA-MLP پیش‌بینی نزدیک بهم داشتند. اما دقت مدل PCA-SVM نسبت به مدل دیگر بالاتر بود. این مطالعه همسو با مطالعه‌ی Wang و همکاران بود و نشان داد که با استفاده از مدل‌های PCA-SVM و PCA-MLP و متغیرهای محیطی به‌دست آمده از داده‌های ماهواره‌ای، می‌توان بروز بروسولوز انسانی را با درجه‌ای از دقت معقول پیش‌بینی کرد (۴).

Wang و همکاران، مطالعه‌ای بر روی رویکرد شبکه‌ی عصبی مصنوعی مبتنی بر داده‌های سنجش از دور جهت پیش‌بینی شیوع بیماری بروسولوز انسانی انجام دادند. در این مطالعه از بروسولوز انسانی به عنوان یک مورد آزمایشی برای شناسایی عوامل مهم محیطی تعیین‌کننده بیماری و پیش‌بینی شیوع آن استفاده شد. یک مدل

به‌طور مشخص در مناطق روستایی و مرتفع‌تر متمرکز است (۱۱). این یافته مشلبه با مطالعه‌ی ما در سبزواری بود که مناطق مرتفع‌تر مانند ششتمد در گروه پرخطر قرار گرفتند. تفاوت‌های اقلیمی و تفاوت در سیاست‌های دامپزشکی و واکسیناسیون می‌تواند علت تفاوت شدت بروز بیماری باشد.

نتیجه‌گیری

در این مطالعه با بهره‌گیری از GIS و دو الگوریتم ترکیبی یادگیری ماشینی، PCA-SVM و PCA-MLP، بروز فضایی بروسلوز انسانی در شهرستان سبزواری مدل‌سازی و پیش‌بینی شد. نتایج نشان داد که مدل PCA-SVM عملکرد بهتری نسبت به PCA-MLP داشته و با دقت و حساسیت بالاتری مناطق پرخطر را شناسایی می‌کند. این پژوهش اثبات کرد که ترکیب روش‌های داده‌محور و هوش مصنوعی همراه با تحلیل فضایی، ابزار قدرتمندی برای شناسایی نواحی پرریسک، مدیریت بیماری‌های زئونوز و برنامه‌ریزی پیشگیرانه خصوصاً در مناطق روستایی کم‌منبع فراهم می‌آورد. برای ارتقاء جامعیت و تعمیم‌پذیری نتایج، پیشنهاد شده است که در مطالعات آینده از سایر الگوریتم‌های یادگیری ماشینی، متغیرهای اجتماعی-اقتصادی، داده‌های زمانی با وضوح بالاتر و گسترش مطالعه به سایر شهرستان‌ها بهره گرفته شود.

تشکر و قدردانی

این مقاله منتج از پایان‌نامه‌ی مقطع دکتری حرفه‌ای رشته‌ی پزشکی با کد ۴۰۱۰۲۳ می‌باشد که در دانشگاه علوم پزشکی سبزواری به تصویب رسیده و با حمایت مالی این دانشگاه به انجام رسیده است. لذا، نویسندگان از معاونت تحقیقات و فناوری دانشگاه علوم پزشکی سبزواری بابت حمایت مالی، و از معاونت بهداشتی دانشگاه بابت در اختیار گذاشتن داده‌های مربوطه تشکر می‌کنند.

به فصل‌ها و تغییرات آب و هوایی و ارتفاع و شیب زمین اشاره کرد که همسو با مطالعه‌ی فوق می‌باشد.

Peng و همکاران در سال ۲۰۲۰ مطالعه‌ای با هدف توزیع مکانی-زمانی بروسلوز انسانی در چین با تجزیه و تحلیل عوامل اجتماعی و محیطی انجام دادند. در این مطالعه تعداد سالانه موارد بروسلوز و نرخ بروز از ۳۱ استان در چین بین سال‌های ۲۰۰۴ تا ۲۰۱۷ به‌دست آمد. همچنین تعداد موارد ماهانه بروسلوز در سال ۲۰۱۸ از مرکز کنترل و پیشگیری از بیماری‌های چین جمع‌آوری شد. نتایج این مطالعه نشان داد که موارد بروسلوز انسانی در سراسر کشور چین گسترش پیدا کرده است. موارد بروسلوز انسانی بیشتر در ماه‌های مارس تا اوت رخ داده و از آوریل تا ژوئیه متمرکز شده است. موجودی دام زنده، تولید ناخالص داخلی، و آب و هوا به طور قابل توجهی با حضور موارد بروسلوز در کشور چین مرتبط بود (۱۰). در مطالعه‌ی حاضر به طور کلی در تمام مناطق در سال‌های ۱۳۹۱، ۱۳۹۳ تا ۱۳۹۶، ۱۳۹۹، ۱۴۰۱ و بیشترین ابتلا در فصل بهار بوده است. در سال‌های ۱۳۹۰، ۱۳۹۲ و ۱۳۹۷ ابتلا در فصول مختلف متغیر بوده است. همچنین در سال‌های ۱۳۹۸ و ۱۴۰۰ بیشترین ابتلا در فصل تابستان بوده است. بنابراین در مطالعه‌ی ما همسو با مطالعه‌ی Peng و همکاران، بیشترین ابتلا در فصل بهار و سپس تابستان بوده است (۱۰).

در مطالعه‌ی Pakzad و همکاران که به بررسی توزیع فضایی بروسلوز در ایران پرداختند، نتایج مشابهی مبنی بر خوشه‌ای بودن توزیع مکانی بیماری در بهار و تابستان گزارش شد. این یافته با نتایج مطالعه‌ی ما تطابق داشت، چرا که الگوی فصلی در هر دو پژوهش بر بروز بالای بیماری در فصل‌های گرم سال تأکید داشت. دلیل این امر می‌تواند افزایش فعالیت دام‌ها، مصرف لبنیات تازه محلی و افزایش تماس‌های انسانی-دامی در فصل‌های گرم باشد (۶).

مطالعه‌ای در چین توسط Liang و همکاران به تحلیل فضایی بروسلوز با استفاده از GIS پرداخت و نشان داد که توزیع بیماری

References

1. Abdullayev R, Kralalik I, Ismayilova R, Ustun N, Talibzade A, Blackburn JK. Analyzing the spatial and temporal distribution of human brucellosis in Azerbaijan (1995-2009) using spatial and spatio-temporal statistics. *BMC Infect Dis* 2012; 12: 1-12.
2. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol* 2010; 140(3-4): 392-8.
3. Sofian M, Aghakhani A, Velayati AA, Banifazl M, Eslamifar A, Ramezani A. Risk factors for human brucellosis in Iran: a case-control study. *Int J Infect Dis* 2008; 12(2): 157-61.
4. Wang J, Jia P, Cuadros DF, Xu M, Wang X, Guo W, et al. A remote sensing data based artificial neural network approach for predicting climate-sensitive infectious disease outbreaks: a case study of human brucellosis. *Remote Sens* 2017; 9(10): 1018.
5. Mollalo A, Mao L, Rashidi P, Glass GE. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. *Int J Environ Res Public Health* 2019; 16(1): 157.
6. Pakzad R, Pakzad I, Safiri S, Shirzadi MR, Mohammadpour M, Behroozi A, et al. Spatiotemporal analysis of brucellosis incidence in Iran from 2011 to 2014 using GIS. *Int J Infect Dis* 2018; 67: 129-36.

7. Tabasi M, Alesheikh AA, Sofizadeh A, Saeidian B, Pradhan B, AlAmri A. A spatio-temporal agent-based approach for modeling the spread of zoonotic cutaneous leishmaniasis in northeast Iran. *Parasites vectors* 2020; 13(1): 727.
8. Rezaeianzadeh M, Tabari H, Arabi Yazdi A, Isik S, Kalin L. Flood flow forecasting using ANN, ANFIS and regression models. *Neural Comput & Applic* 2014; 25(1): 25-37.
9. Mohammadinia A, Saeidian B, Pradhan B, Ghaemi Z. Prediction mapping of human leptospirosis using ANN, GWR, SVM and GLM approaches. *BMC Infect Dis* 2019; 19(1): 971.
10. Peng C, Li Y-J, Huang D-S, Guan P. Spatial-temporal distribution of human brucellosis in mainland China from 2004 to 2017 and an analysis of social and environmental factors. *Environ Health Prev Med* 2020; 25(1): 1-14.
11. Liang P, Zhao Y, Zhao J, Pan D, Guo Z. The spatiotemporal distribution of human brucellosis in mainland China from 2007-2016. *BMC Infect Dis* 2020; 20(1): 249.

Spatial Modeling and Prediction of Human Brucellosis in Sabzevar Using Machine Learning: A Comparison of PCA MLP and PCA SVM

Ayoob Rastegar¹, Ali Oghazyan², Shima Amiri Parsa³, Ali Pouryousef⁴,
Mohammad-Shafi Mojadadi⁴

Original Article

Abstract

Background: Brucellosis (Malta fever) is a zoonotic bacterial disease affecting both humans and animals, prevalent in various regions of Iran, including Sabzevar County. This study aimed to model and predict the spatial incidence of human brucellosis using Geographic Information Systems (GIS) and machine learning algorithms, specifically Principal Component Analysis-Multilayer Perceptron (PCA-MLP) and Principal Component Analysis-Support Vector Machine (PCA-SVM), within this county.

Methods: Data related to brucellosis incidence, population information, climate, vegetation cover, and topography were collected for the period between 2011 and 2022 (1390 to 1401 Iranian calendar years). The data were processed using GIS and subsequently employed with the PCA-MLP and PCA-SVM machine learning algorithms to model and predict the spatial distribution of the disease.

Findings: In this study, seasonal analysis revealed that the peak incidence of brucellosis occurred in spring during the years 2012, 2014–2017, 2020, and 2022. However, in 2019 and 2021, this peak shifted to summer, and in some years (2011, 2013, and 2018), the seasonal distribution of the disease showed an unstable trend. Furthermore, air humidity and altitude above sea level were identified as the most significant predictors of incidence. The PCA-MLP model predicted areas like Sheshatmad and Dastooran as having the highest number of human cases, while the PCA-SVM model, in addition to these, also categorized some other studied areas into the high-risk group.

Conclusion: The PCA-SVM algorithm demonstrated more accurate performance in modeling and predicting disease incidence compared to PCA-MLP, more precisely identifying high-risk areas such as Sheshatmad and Dastooran. The findings of this study can assist health managers in allocating resources and designing preventive interventions in high-risk regions.

Keywords: Brucellosis, Geographic information systems, Machine learning algorithms, Neural network models, Principal component analysis, Prediction

Citation: Rastegar A, Oghazyan A, Amiri Parsa Sh, Pouryousef A, Mojadadi M. **Spatial Modeling and Prediction of Human Brucellosis in Sabzevar Using Machine Learning: A Comparison of PCA MLP and PCA SVM.** J Isfahan Med Sch 2026; 44(854): 348- 57.

1- Associate Professor, Department of Environmental Health Engineering, School of Health, North Khorasan University of Medical Sciences, Bojnourd, Iran

2- PhD Student, Department of Environmental Health Engineering, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Medical Student, Student Research Committee, Sabzevar University of Medical Sciences, Sabzevar, Iran

4- Assistant Professor, Leishmaniasis Research Center, Sabzevar University of Medical Sciences, Sabzevar, Iran

5- Professor, Leishmaniasis Research Center, Sabzevar University of Medical Sciences, Sabzevar, Iran

Corresponding Author: Mohammad-Shafi Mojadadi, Professor, Leishmaniasis Research Center, Sabzevar University of Medical Sciences, Sabzevar, Iran; Email: mojadadi@gmail.com