

## پیش‌بینی تعداد موارد بروسلوز براساس پارامترهای اقلیمی با استفاده از روش‌های داده‌کاوی شبکه‌های عصبی مصنوعی پرسپترون چند لایه، تابع پایه شعاعی و نزدیک‌ترین همسایگی

لیلی تاپاک<sup>۱</sup>، نسرین شیرمحمدی خرم<sup>۲</sup>، امید حمیدی<sup>۳</sup>، زهره مریانجی<sup>۴</sup>

<sup>۱</sup> استادیار، گروه آمار زیستی، دانشکده بهداشت و مرکز تحقیقات مدل‌سازی بیماری‌های غیر واگیر، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۲</sup> کارشناس ارشد آمار زیستی، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۳</sup> مربی، گروه علوم پایه، دانشگاه صنعتی همدان، همدان، ایران

<sup>۴</sup> استادیار، گروه جغرافیا، دانشگاه سیدجمال‌الدین اسدآبادی، اسدآباد، ایران

نویسنده رابط: نسرین شیرمحمدی خرم، نشانی: همدان، دانشگاه علوم پزشکی همدان، دانشکده بهداشت، گروه آمار زیستی همدان، ایران. کدپستی: ۶۵۱۷۶۶۹۶۶۴. تلفن:

۹۸۸۱۳۸۳۸۰۰۲۵ پست الکترونیک: n.shirmohammadi@edu.umsha.ac.ir

تاریخ دریافت: ۹۶/۶/۲۶؛ پذیرش: ۹۶/۱۲/۱۲

**مقدمه و اهداف:** شناسایی مدل‌های آماری دارای پیش‌بینی‌های دقیق در تعیین دقیق و به‌هنگام طغیان بیماری‌های عفونی در نظام مراقبت بهداشتی این بیماری‌ها بسیار با اهمیت است. این مطالعه با هدف ارزیابی و مقایسه عملکرد سه روش یادگیری ماشین در مدل‌سازی و پیش‌بینی سری زمانی بروسلوز بر اساس پارامترهای اقلیمی انجام شد.

**روش کار:** در این مطالعه موارد بروسلوز انسانی و پارامترهای اقلیمی به‌صورت ماهانه، در طول ۱۲ سال (۹۵-۱۳۸۳) از استان همدان واقع در غرب ایران تحلیل شد. داده‌ها به دو زیرمجموعه آموزش (۸۰ درصد) و آزمون (۲۰ درصد) تقسیم شد. روش‌های تابع پایه شعاعی و چند لایه پرسپترون و نزدیک‌ترین همسایه سری زمانی به هر زیرمجموعه برازش شد. ارزیابی عملکرد مدل‌ها با استفاده از معیارهای RMSE، MAE، MRAE، R2 و ICC انجام شد.

**یافته‌ها:** نتایج نشان داد که مقادیر معیارهای (۲۳/۷۹) RMSE، (۲۰/۵۶) MAE، (۰/۲۵) MRAE برای مدل شبکه عصبی چند لایه پرسپترون کوچک‌تر از مقادیر آن‌ها در دو مدل دیگر بود. هم‌چنین، در این مدل مقادیر بزرگ‌تری برای معیارهای (۰/۶۱) R2 و (۰/۷۵) ICC به‌دست آمد. بنابراین مدل شبکه‌ی عصبی چند لایه پرسپترون در پیش‌بینی داده‌های مورد مطالعه عملکرد بهتری داشت. دما نسبت به سایر پارامترهای اقلیمی مؤثرترین عامل در بروز این بیماری بود.

**نتیجه‌گیری:** شبکه عصبی چندلایه پرسپترون می‌تواند به‌عنوان یک روش کارا برای تشخیص رفتار روند بروسلوز در طول زمان به کار رود. با این حال مطالعات بیش‌تری با هدف کاربرد و مقایسه این روش‌ها برای شناسایی مناسب‌ترین روش پیش‌بینی روند این بیماری مورد نیاز است.

**واژگان کلیدی:** بروسلوز انسانی، شبکه عصبی مصنوعی، نزدیک‌ترین همسایه، پیش‌بینی

### مقدمه

(۳-۶). این بیماری اثرات بهداشتی و اقتصادی شدیدی دارد (۴،۵). به دلیل تنوع نشانه‌های این بیماری در انسان و مشکلاتی که در تشخیص این بیماری وجود دارد، شیوع آن در جهان کم‌تر از شیوع واقعی آن گزارش می‌شود (۷). بر اساس گزارش سازمان جهانی بهداشت سالانه حدود ۵۰۰۰۰۰ مورد جدید بروسلوز در جهان گزارش می‌شود که بیش‌تر آن مربوط به کشورهای در حال توسعه است (۸). باوجود این که بروسلوز در بسیاری از کشورهای صنعتی حذف شده است، این بیماری هنوز یک تهدید بهداشتی جدی در برخی از کشورها مانند ایران است (۳،۴،۹). بروسلوز انسانی در ایران به‌عنوان یک بیماری اندمیک شناخته می‌شود.

رخداد بیماری‌های عفونی اغلب دارای یک الگوی منظم و چرخه‌ای است. تعیین دقیق و به‌هنگام تغییرات رفتار روند بیماری‌های عفونی در طول زمان در مدیریت و کنترل این نوع بیماری‌ها، امری ضروری است (۱،۲). بروسلوز یک بیماری عفونی شایع و قابل انتقال بین انسان و حیوان در جهان است که می‌تواند به‌صورت غیر مستقیم و با مصرف محصولات لبنی غیر پاستوریزه یا با تماس مستقیم با حامل باکتری بروسلا از دام به انسان انتقال یابد (۳-۵). افراد آلوده به باکتری بروسلا اغلب از شرایطی مشابه شرایط یک بیماری ناتوان‌کننده و مزمن مانند تب، درد مفاصل، تعریق، خستگی و افسردگی برای هفته‌ها و ماه‌ها رنج می‌برند

و ساختار شبکه‌ای ساده‌تر در مقایسه با دیگر شبکه‌های عصبی را دارا است. در حالی که، MLP قابلیت تعمیم‌پذیری بهتری دارد (۲۳). عملکرد این روش در پیش‌بینی نسبت به روش‌های کلاسیک بهتر است (۱۹). یکی دیگر از روش‌های یادگیری ماشین، روش نزدیکترین همسایگی ( $NN^1$ ) است (۲۴) که یک روش ناپارامتری است و در مدل‌بندی فرایندهای غیر خطی در حوزه‌های مختلف علوم به کار رفته است و عملکرد خوبی در رگرسیون و مسائل طبقه‌بندی نشان داده است. این روش مبتنی بر الگوشناسی و استفاده از اصل تشابه و نزدیکی داده‌ها است. چنین رویکردی در برآورد ویژگی‌های موردنظر، هنگامی که نوع رابطه بین داده‌های ورودی و خروجی روشن نباشد، از سودمندی بیش‌تری برخوردار است. پژوهشگران بسیاری عملکرد مناسب این روش و مزایای آن را در مسائل پیش‌بینی تأیید کرده‌اند (۲۲، ۲۴)، اما با وجود کاربرد وسیع و عملکرد مناسب، تا کنون عملکرد این دو روش در داده‌های بروسولوز مورد ارزیابی قرار نگرفته است.

با توجه به اهمیت پیش‌بینی دقیق و به‌هنگام تغییرات روند بیماری در مدیریت طغیان بیماری‌ها، پیش‌بینی دقیق با استفاده از روش‌های آماری ضروری است. هم‌چنین، با توجه به این که عملکرد روش‌های مختلف وابسته به داده‌هاست و این امکان وجود دارد که در داده‌های مختلف، عملکرد روش‌ها متفاوت باشد، بنابراین هدف این مطالعه ارزیابی عملکرد روش‌های شبکه‌های عصبی مصنوعی (تابع پایه شعاعی و چند لایه پرسپترون هر کدام به‌طور جداگانه) و نزدیک‌ترین همسایگی در پیش‌بینی روند بروسولوز انسانی و شناسایی مدلی با پیش‌بینی بهتر و استفاده از آن در سامانه بهداشت عمومی در کنترل و پیش‌گیری بروز بالای این بیماری است. هم‌چنین از دیدگاه اقلیمی، تعیین شرایط و وضعیت اپیدمیولوژی بروسولوز با توجه به شرایط محیطی در مناطق مختلف ضروری به نظر می‌رسد. این امر مستلزم بررسی و مطالعه ویژگی‌های محیطی هر منطقه است. از مهم‌ترین این ویژگی‌ها می‌توان به شرایط اقلیمی و آب و هوایی منطقه اشاره کرد. با توجه به این که عامل بروسولوز، باکتریایی است بنابراین شناسایی شرایط محیطی مناسب برای رشد این باکتری مانند شرایط دمایی، رطوبت، بارش، باد می‌تواند در مدیریت و کنترل این بیماری کمک شایانی نماید. هدف دیگر این پژوهش تعیین اثرات پارامترهای اقلیمی مانند دما، بارش، حداقل، حداکثر دما، رطوبت، باد، ساعت‌های آفتابی و فشار هوا بر این بیماری با

رخداد سالانه این بیماری در کشور ایران به‌ویژه در غرب و شمال‌غرب بالا گزارش شده است که این موضوع منجر به مشکلات گسترده‌ای خواهد شد (۴). بنابراین برنامه‌ریزی راهبردی و کنترل مبتنی بر مدیریت کاربردی برای توسعه و بهبود بهداشتی این بیماری ضروری است.

برای رسیدن به این هدف، داده‌های با کیفیت بالا و روش‌های پیش‌بینی با کم‌ترین خطا مورد نیاز است (۱۰). روش‌های پیش‌بینی آماری با پیش‌بینی دقیق رخدادهای بیماری و دارای عملکرد خوب یک ابزار مفید در پیش‌بینی تغییرات روند بیماری‌ها در آینده هستند (۲). در مطالعه‌های مختلف روش‌های آماری متفاوتی برای پیش‌بینی میزان بروز بیماری‌های عفونی استفاده شده‌اند. مدل‌های سری زمانی مختلف از جمله روش هموارسازی (۱)، روش تجزیه (۱۱)، مدل‌های سری زمانی چندسطحی (۱۲)، مدل‌های میانگین متحرک اتورگرسیو (۱۷-۱۳)، به‌طور گسترده‌ای در پیش‌بینی بیماری‌های عفونی به‌کار گرفته شده‌اند. مدل‌های سری زمانی سعی بر پیش‌بینی رفتارهای اپیدمیولوژی با مدل‌سازی داده‌های مراقبت تاریخی دارند (۲). استفاده از این مدل‌ها نیازمند تعیین دقیق پارامترهای ریاضی برای توصیف پدیده‌های مختلف و برقراری فرضیه‌های زیربنایی مدل به‌ویژه خطی بودن رابطه رگرسیونی است (۱).

در سال‌های اخیر، مدل‌های سری زمانی مبتنی بر روش‌های یادگیری ماشین در مدل‌سازی تغییرات میزان بروز بیماری‌های عفونی مورد استفاده و توجه از سوی پژوهشگران قرار گرفته‌اند. از جمله این روش‌ها شبکه‌های عصبی مصنوعی ( $ANN^1$ ) است که به‌منظور پیش‌بینی در علوم مختلف به کار برده شده‌اند (۲۱-۱۸) و نشان داده شده است که این روش به‌طور مؤثری عملکرد پیش‌بینی بهتری نسبت به روش‌های کلاسیک دارد. مدل شبکه عصبی مصنوعی یک روش توانمند غیرخطی در مدل‌سازی داده‌ها است که می‌تواند به‌خوبی روابط پیچیده بین متغیرهای پیش‌گو و هدف را بدون در نظر گرفتن هیچ فرضیه اولیه و دانش قبلی از روابط بین پارامترهای مورد مطالعه، مدل‌سازی کند (۲۲). دو روش پیشرو در مبحث شبکه‌های عصبی، شبکه‌ی تابع پایه شعاعی ( $RBF^2$ ) و شبکه پرسپترون چند لایه ( $MLP^3$ ) است.  $RBF$  نوع رایج‌تری از یادگیری شبکه عصبی است و تنها به بخش محدودی از فضای ورودی پاسخ می‌دهد و روش آموزشی سریع‌تر و دقیق‌تر

<sup>1</sup>Artificial Neural Network

<sup>2</sup>Radial Basis Function

<sup>3</sup>Multilayer Perceptron

<sup>۴</sup>Nearest Neighborhood

سنوپتیک همدان به دست آمد. به منظور بررسی اعتبار مدل‌های به‌کار رفته در این مطالعه، داده‌های ماهانه به دو دسته آموزش و آزمون تقسیم شدند. این تقسیم‌بندی بر اساس چگونگی ارزیابی عملکرد در داده‌های سری زمانی انجام شد. در مطالعه‌های انجام شده روی داده‌های سری‌های زمانی، نسبت ۷۰ یا ۸۰ درصد داده‌ها از ابتدای سری به عنوان مجموعه آموزشی و باقی‌مانده داده‌ها تا انتهای سری به عنوان مجموعه آزمون در نظر گرفته می‌شود (۲۶، ۲۷). بنابراین در این مطالعه نیز نسبت ۸۰ درصد از داده‌ها به عنوان مجموعه آموزش (از فروردین ۱۳۸۳ تا آبان ۱۳۹۳) و ۲۰ درصد باقی‌مانده (آذر ۱۳۹۳ تا اسفند ۱۳۹۵) به عنوان مجموعه آزمون در نظر گرفته شد.

#### مدل‌ها

در این مطالعه، ۳ روش شبکه عصبی مصنوعی تابع پایه شعاعی و چند لایه پرسپترون و روش نزدیک‌ترین همسایگی سری زمانی برای پیش‌بینی تغییرات ماهانه فراوانی بروسلوز بر اساس پارامترهای اقلیمی استفاده شد. خودهمبستگی‌ها با تأخیر ۱-۱۲ ماه برای داده‌های ماهانه بروسلوز نیز در مدل وارد شد. سطح معنی‌داری ۰/۰۵ در نظر گرفته شد.

#### K- نزدیک‌ترین همسایگی

این روش یک تکنیک ناپارامتری است که دارای الگوریتم ساده‌ای است. در این روش، یک گروه شامل K رکورد از مجموعه رکوردهای آموزشی که نزدیک‌ترین رکوردها به رکورد آزمایشی باشند، انتخاب شده و متغیر پاسخ کمی با استفاده از برخی معیارهای تشابه مانند توابع فاصله معمول پیش‌بینی می‌شود (۲۶). مزیت اصلی این مدل ناپارامتری، اجرای ساده، قابلیت مدل‌سازی غیر خطی و عملکرد با بازدهی بالا در برخورد با تعداد دسته‌های زیاد از داده‌ها است (۲۷). رگرسیون K نزدیک‌ترین همسایگی یک ابزار نیرومند تحلیل داده است که در این روش استفاده از تعداد همسایگی‌های بیش‌تر منجر به برآوردهای هموار مرزهای دسته‌بندی در مسائل رده‌بندی و برآوردهای نزدیک به هم و با دقت کم‌تر در رگرسیون می‌شود، در حالی که تعداد کم‌تر همسایگی‌ها، برآوردهای ناریب را نتیجه می‌دهد و منجر به مسأله بیش‌برازشی می‌شود. از طرفی افزایش تعداد همسایگی‌ها منجر به بهبود صحت برآوردهای به دست آمده می‌شود. برآوردگر نزدیک‌ترین همسایگی با استفاده از مقدار میانگین تعداد ثابتی از مشاهده‌های نزدیک و همسایگی به دست می‌آید (۲۶، ۲۷).

استفاده از روش‌های یادگیری ماشین است. یافته‌های این مطالعه می‌تواند با تعیین مناسب‌ترین روش در کنترل و پیشگیری طغیان‌ها برای اپیدمیولوژیست‌ها مفید واقع شود.

## روش کار

### داده‌ها و منطقه مورد مطالعه

این مطالعه روی داده‌های سری زمانی تعداد موارد بروسلوز و پارامترهای اقلیمی شامل میانگین و حداقل و حداکثر دما، رطوبت نسبی، باد، ساعت‌های آفتابی، فشار هوا به‌صورت ماهانه در استان همدان انجام شد. استان همدان در غرب کشور و در بین رشته کوه زاگرس واقع شده است و به علت ارتفاع بالا از سطح دریا به‌عنوان یکی از استان‌های سردسیری کشور محسوب می‌شود. آب و هوای همدان نیز مانند سایر ناحیه‌های غرب کشور تحت کنترل عوامل بیرونی قرار دارد. همدان از مناطق کوهستانی، مرتفع، سرد و بادخیز کشور محسوب می‌شود. به دلیل شرایط جغرافیایی و توپوگرافی دارای آب و هوای نسبتاً متغیر است. بر اساس طبقه‌بندی اقلیمی بلور و دمارتن و آمبرژه اقلیم همدان جزء اقلیم نیمه‌خشک محسوب می‌شود. بر اساس آخرین تقسیمات کشوری توسط وزارت کشور در سال ۲۰۱۳ میلادی، این استان دارای مساحت ۱۹۵۴۶ کیلومتر مربع با ۹ شهرستان همدان، تویسرکان، نهاوند، ملایر، اسدآباد، بهار، فامنین، رزن و کبودرآهنگ است.

داده‌های بالینی و اپیدمیولوژی بیماران بر اساس راهنمای کشوری، به صورت آنلاین در سامانه مراقبت بهداشتی ثبت می‌شود. بر اساس این راهنما مواردی که دارای علائم بالینی اپیدمیولوژی بروسلوز مانند تب، شب‌اداری و علائم پاراکلینیکی (نتیجه دو آزمون آزمایشگاهی متداول برای تشخیص بروسلوز) شامل رایت<sup>۱</sup> (آزمون تشخیصی بروسلوز که مقادیر بزرگ‌تر از ۱/۸۰ نشان دهنده مثبت بیماری است) و 2ME<sup>۲</sup> (آزمون تأییدی بروسلوز که در صورتی که مقدار آن بیش‌تر یا مساوی ۱/۴۰ باشد نشان دهنده مثبت بودن بیماری است) باشند، بیمار محسوب می‌شوند (۶، ۲۵).

در این مطالعه، روند تعداد موارد بروسلوز انسانی بر اساس روند پارامترهای اقلیمی ماهانه طی سال‌های ۹۵-۱۳۸۳ در استان همدان تحلیل شد. داده‌های تعداد موارد بروسلوز از سامانه‌ی اطلاعات معاونت بهداشتی دانشگاه علوم پزشکی استان همدان استخراج شد و داده‌های مربوط به پارامترهای اقلیمی از ایستگاه

<sup>۱</sup>Wright

<sup>۲</sup> Mercaptoethanol Brucella agglutination test

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_{obs} - Y_{pred})^2}$$

$$MAE = \frac{1}{n} \sum |Y_{obs} - Y_{pred}|$$

$$R^2 = 1 - \frac{\sum (Y_{obs} - MY_{pred})^2}{\sum (Y_{obs} - \bar{Y})^2}$$

$$MARE = \frac{1}{n} \sum \frac{|Y_{obs} - Y_{pred}|}{Y_{obs}}$$

$$ICC = \frac{MS_B}{MS_B + MS_E}$$

در روابط بالا  $Y_{obs}$  و  $Y_{pred}$  به ترتیب نشان دهنده مقادیر مشاهده شده و پیش‌بینی شده تعداد موارد بروسولوز هستند.  $MS_B$  تغییرات بین مشاهده‌ها و  $MS_E$  تغییرات درون مشاهده‌ها است.

### نرم‌افزارها

در این مطالعه روش‌های مورد استفاده روی مجموعه داده‌ها با استفاده از نرم‌افزار SPSS نسخه ۲۳ انجام شد. برای برازش مدل‌های MLP، ۲-۱ لایه پنهان و توابع مختلف هایبربولک تانژانت و سیگموئید در لایه‌ی پنهان و توابع یکنواخت، هایبربولیک تانژانت و سیگموئید در لایه‌ی خروجی در نظر گرفته شد. در برازش مدل‌ها متغیرهای کمکی و متغیر وابسته نرمال‌سازی شدند که این کار در با استفاده از رابطه زیر انجام شد (۲۸):

$$Y_{Normalized} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}}$$

### یافته‌ها

جدول شماره ۱ میانگین، انحراف استاندارد، حداقل و حداکثر تعداد موارد بروسولوز را برای کل داده‌ها و به تفکیک مجموعه داده‌های آموزش و آزمایش نشان می‌دهد. طی سال‌های مورد مطالعه کم‌ترین و بیش‌ترین تعداد موارد بروسولوز به ترتیب با مقادیر ۴ و ۳۲۳ مربوط به اسفندماه سال ۱۳۸۷ و تیرماه سال ۱۳۸۴ بود. شکل شماره ۱ روند این بیماری را طی سال‌های مورد مطالعه در استان همدان نشان می‌دهد. بیش‌ترین تعداد موارد بروسولوز در سال ۱۳۸۴ رخ داد. بیش‌ترین تعداد موارد بروسولوز فصلی با میانگین (انحراف معیار) ۱۲۸/۵ (۸۴/۰۲) به فصل تابستان و تیرماه و کم‌ترین آن با میانگین (انحراف معیار) ۶۹/۹۴ (۳۵/۷۲) به فصل زمستان اختصاص داشت.

تحلیل همبستگی برای انتخاب ورودی‌های مناسب در مدل‌سازی انجام شد و ضرایب خودهمبستگی معنی‌دار به عنوان

### شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی یک ابزار ریاضی برای پردازش اطلاعاتی است که در زمینه‌های زیادی از علوم به‌منظور پیش‌بینی، تشخیص الگو و مسأله رده‌بندی به کار گرفته شده‌اند (۲۶). چندین نود (Node) ورودی و وزن‌هایی که نودها را به یکدیگر مرتبط می‌کنند در یک مدل شبکه عصبی وجود دارد. انواع شبکه‌های عصبی وجود دارد که از میان آن‌ها دو روش پرسپترون چند لایه و تابع شعاعی پایه در مطالعه‌های زیادی مورد استفاده قرار گرفته و با یکدیگر مقایسه شده است. مدل پرسپترون چند لایه یک نوع خاص از این روش است که دارای تابع فعال‌سازی غیر خطی مانند سیگموئید در لایه‌ی پنهان و تابع خطی در لایه خروجی است (۲۶). در این مدل رابطه بین لایه ورودی و مخفی به صورت خطی است و فرم تابعی این مدل به صورت زیر است:

$$y_j = f \left( \sum_{i=1}^N w_{ji} x_i + b_j \right)$$

که در آن  $x$  مقدار نودال در لایه‌ی پیشین،  $y$  مقدار نودال در لایه حاضر،  $b$  عرض از مبدأ یا اربیبی در لایه‌ی حاضر و  $w$  ضرایب رگرسیونی یا وزن‌ها هستند (۲۹، ۲۸). در حالی که در مدل تابع شعاعی پایه که همانند مدل پرسپترون چندلایه شامل ۳ لایه ورودی، مخفی و خروجی است، رابطه‌ی بین لایه‌ی مخفی و لایه‌ی ورودی به صورت یک تابع غیر خطی بر اساس توابع فعال‌سازی شعاعی پایه است (۳۰).

### معیارهای ارزیابی

برای ارزیابی و مقایسه دقت پیش‌بینی و عملکرد مدل‌های مورد نظر در مدل‌سازی داده‌های سری زمانی مورد مطالعه از معیارهای مجذور میانگین مربعات خطا (RMSE<sup>۱</sup>)، میانگین قدر مطلق خطا (MAE<sup>۲</sup>)، میانگین نسبی قدر مطلق خطا (MARE<sup>۳</sup>)، ضریب تعیین (R<sup>۲</sup>) و ضریب همبستگی درون کلاسی (ICC<sup>۴</sup>) استفاده شد که با استفاده از روابط زیر محاسبه می‌شود (۲۳، ۲۸):

<sup>۱</sup> Root Mean Square Error  
<sup>۲</sup> Mean Absolute Error  
<sup>۳</sup> Mean Absolute Root Error  
<sup>۴</sup> Intraclass Correlation Coefficient

به دست داد، بنابراین عملکرد روش شبکه عصبی مصنوعی MLP در مقایسه با RBF و روش K نزدیک‌ترین همسایگی بهتر بود. به علاوه شکل ۴ مقادیر مشاهده شده و برآوردهای تعداد موارد بروسلوز (پیش‌بینی شده) حاصل از ۳ روش را با استفاده از نمودار پراکندگی در مقابل یکدیگر نمایش می‌دهد. همان‌طور که ملاحظه می‌شود، همه نقاط در یک چهارم نخست افتادند که نشان می‌دهد علامت مقادیر برآورد شده با مقادیر مشاهده شده یکسان است. همچنین مقدار معنی‌داری مدل رگرسیونی برازش شده محاسبه شد که برای هر ۳ روش RBF، MLP و KNN کوچک‌تر از ۰/۰۰۱ به دست آمد که نشان دهنده معنی‌دار بودن و اعتبار توافق بین مقادیر مشاهده شده و مقادیر پیش‌بینی شده در هر ۳ مدل است. مقدار شیب نزدیک به عدد ۱ نشان‌دهنده عملکرد بهتر آن روش است. با توجه به یافته‌های شکل شماره ۴، شیب خط رگرسیونی برای مدل MLP نسبت به ۲ روش دیگر به ۱ نزدیک‌تر بود که نشان دهنده عملکرد بهتر این روش است.

بر اساس یافته‌های به‌دست آمده، به‌طور کلی روش شبکه عصبی مصنوعی MLP در دوره‌ی زمانی مورد مطالعه در مدل‌سازی تغییرات ماهانه بروسلوز بر اساس پارامترهای اقلیمی عملکرد بهتری داشت.

اهمیت متغیرهای مورد استفاده در روش پرسپترون چندلایه در نمودار شکل شماره ۵ نشان داده شده است. همان‌طور که از شکل مشهود است، بیش‌تر متغیرهای اقلیمی مانند متغیرهای دما و تعداد ساعت‌های تابش آفتاب از جمله متغیرهای تأثیرگذار در تعداد موارد بروسلوز بودند.

ورودی‌ها در نظر گرفته شدند. دو روش شبکه عصبی مصنوعی و روش K نزدیک‌ترین همسایگی به داده‌ها برازش شد.

به منظور مقایسه عملکرد ۳ روش، معیارهای RMSE، MAE، MARE، R<sup>2</sup>، ICC در مجموعه‌های آموزش و آزمایش محاسبه شد. یافته‌های مجموعه آزمایش در جدول شماره ۲ ارائه شده است. با توجه به این یافته‌ها، مقادیر RMSE، MAE، MARE برای روش شبکه عصبی مصنوعی MLP در مقایسه با RBF و مدل K نزدیک‌ترین همسایگی مقادیر کوچک‌تری به دست دادند. از طرف دیگر مقادیر ICC، R<sup>2</sup> در روش شبکه عصبی مصنوعی MLP در مقایسه با روش RBF و K نزدیک‌ترین همسایگی، به یک نزدیک‌تر بود. این یافته‌ها به وضوح نشان می‌دهند که روش شبکه‌ی عصبی مصنوعی MLP در مقایسه با دو روش دیگر در مدل‌سازی و پیش‌بینی برای مجموعه داده سری زمانی بروسلوز ماهانه استان همدان بر اساس پارامترهای اقلیمی این استان عملکرد بهتری داشت.

تغییرات زمانی تعداد موارد مشاهده شده بروسلوز و مقادیر برآورد شده از دو روش شبکه عصبی مصنوعی و K نزدیک‌ترین همسایگی برای مجموعه دوره‌ی آزمایش در شکل ۲ ترسیم شده است. همان‌طور که در شکل شماره ۲ دیده می‌شود در ماه‌های فصل بهار فراوانی‌های بروسلوز افزایش داشته است. همچنین این شکل نشان می‌دهد که مقادیر پیش‌بینی شده از روش شبکه عصبی مصنوعی MLP بهتر از RBF و روش K نزدیک‌ترین همسایگی است.

نمودارهای باقی‌مانده ۲ روش نیز در شکل شماره ۳ به تصویر کشیده شد. روش شبکه عصبی MLP باقی‌مانده‌های کوچک‌تری را

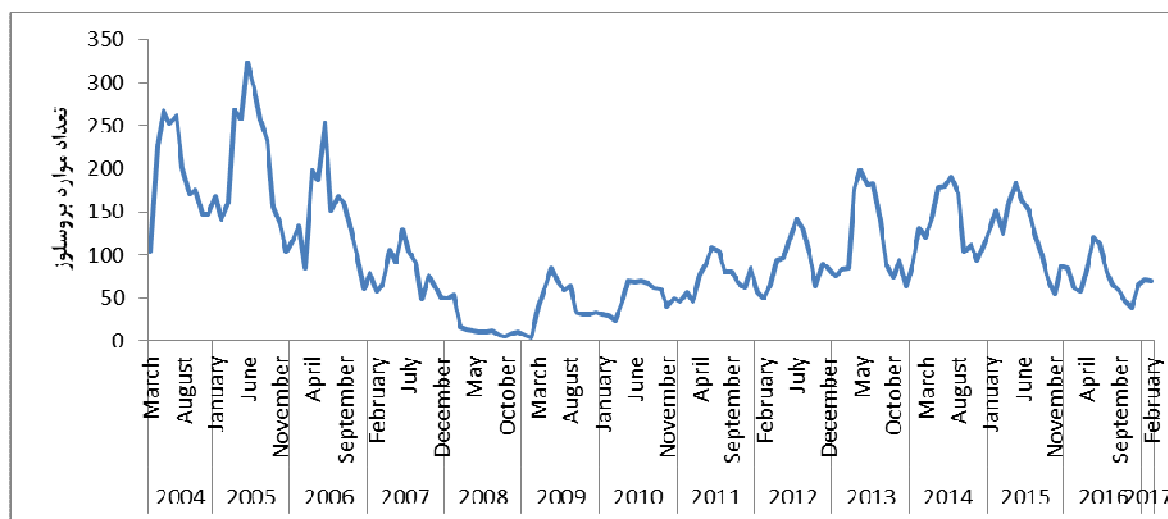
جدول شماره ۱- آماره‌های توصیفی داده‌های سری زمانی تعداد موارد بروسلوز و پارامترهای اقلیمی.

| متغیر               | مجموعه        | حداقل  | حداکثر | میانگین | انحراف استاندارد |
|---------------------|---------------|--------|--------|---------|------------------|
| کل داده‌ها          |               | ۴      | ۳۲۳    | ۹۵/۹۰   | ۶۱/۸۴            |
| تعداد موارد بروسلوز | مجموعه آموزش  | ۴      | ۳۲۳    | ۹۵/۵۵   | ۶۶/۲۰            |
|                     | مجموعه آزمایش | ۳۹     | ۱۸۴    | ۹۹/۴۲   | ۳۹/۸۴            |
| متوسط دمای بیشینه   | کل داده‌ها    | -۱/۲۳  | ۳۷/۰۳  | ۲۰/۵۹   | ۱۰/۵۴            |
|                     | مجموعه آموزش  | -۱/۲۳  | ۳۷/۰۳  | ۲۰/۸۹   | ۱۰/۵۳            |
|                     | مجموعه آزمایش | ۳/۰۹   | ۳۵/۹۸  | ۱۹/۳۶   | ۱۰/۶۸            |
| متوسط دمای کمینه    | کل داده‌ها    | -۱۲/۹۷ | ۱۸/۱۱  | ۵/۶۲    | ۷/۵۲             |
|                     | مجموعه آموزش  | -۱۲/۹۷ | ۱۷/۸۰  | ۵/۹۶    | ۷/۴۹             |
|                     | مجموعه آزمایش | -۶/۱۷  | ۱۸/۱۱  | ۴/۲۳    | ۷/۶۴             |
| بارش ماهانه         | کل داده‌ها    | ۰      | ۱۲۳/۱۵ | ۲۹/۱۵   | ۲۸/۶۱            |
|                     | مجموعه آموزش  | ۰      | ۱۱۰/۹۲ | ۲۸/۷۷   | ۲۷/۸۹            |

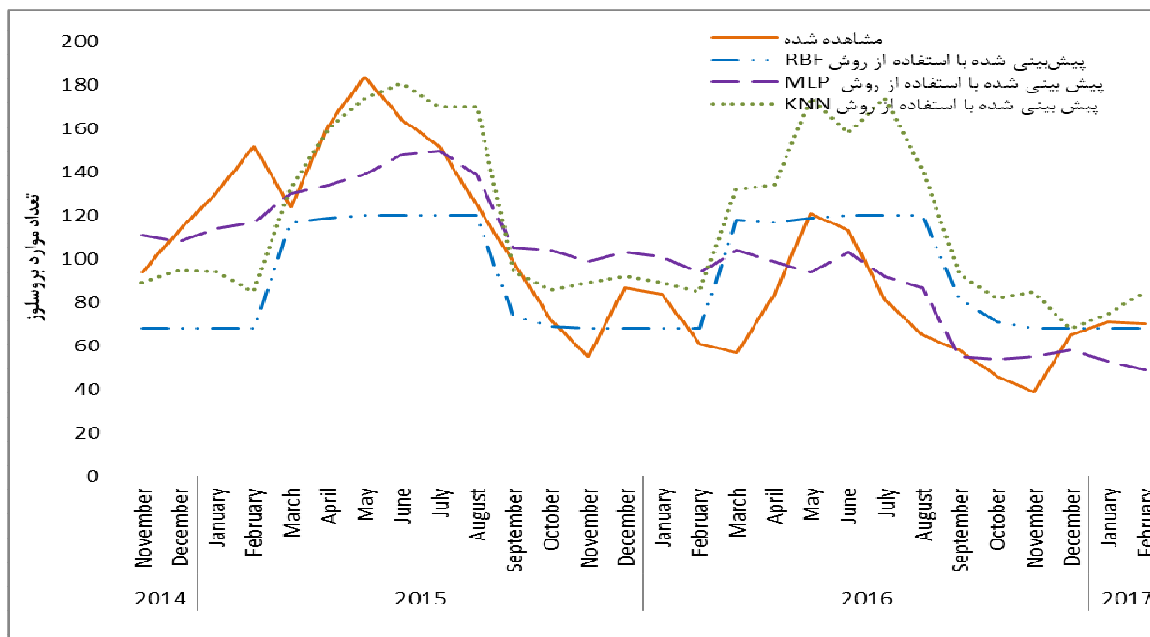
|       |        |        |        |               |                  |
|-------|--------|--------|--------|---------------|------------------|
| ۳۱/۹۶ | ۳۰/۷۳  | ۱۲۳/۱۵ | ۰      | مجموعه آزمایش |                  |
| ۶۸/۱۹ | ۲۵۵/۹۴ | ۳۸۲/۱۱ | ۱۳۴/۲۷ | کل داده‌ها    |                  |
| ۶۸/۵۶ | ۲۵۷/۴۶ | ۳۸۲/۱۱ | ۱۳۴/۲۷ | مجموعه آموزش  | ساعت آفتابی      |
| ۶۷/۴۶ | ۲۴۹/۵۸ | ۳۶۳/۶۳ | ۱۳۶/۸۷ | مجموعه آزمایش |                  |
| ۹/۶۰  | ۱۳/۵۴  | ۲۸/۸۱  | -۷/۸۷  | کل داده‌ها    |                  |
| ۹/۵۱  | ۱۳/۷۲  | ۲۸/۶۶  | -۷/۸۷  | مجموعه آموزش  | میانگین دما      |
| ۱۰/۱۱ | ۱۲/۷۶  | ۲۸/۸۱  | -۱/۱۷  | مجموعه آزمایش |                  |
| ۳/۲۶  | ۸۲۷/۵۸ | ۸۳۴/۶۵ | ۸۲۰/۴۱ | کل داده‌ها    |                  |
| ۳/۱۱  | ۸۲۶/۸۸ | ۸۳۴/۲۴ | ۸۲۰/۴۱ | مجموعه آموزش  | فشار هوا         |
| ۲/۱۴  | ۸۳۰/۴۷ | ۸۳۴/۶۵ | ۸۲۵/۸۹ | مجموعه آزمایش |                  |
| ۱۷/۰۴ | ۴۶/۳۲  | ۸۱/۱۱  | ۲۰/۲۳  | کل داده‌ها    |                  |
| ۱۶/۸۹ | ۴۶/۱۴  | ۸۰/۰۱  | ۲۱/۲۴  | مجموعه آموزش  | میانگین رطوبت    |
| ۱۷/۹۶ | ۴۷/۰۶  | ۸۱/۱۱  | ۲۰/۲۳  | مجموعه آزمایش |                  |
| ۰/۵۷  | ۲/۴۶   | ۴/۲۴   | ۱/۱۶   | کل داده‌ها    |                  |
| ۰/۵۷  | ۲/۴۲   | ۴/۲۴   | ۱/۴۴   | مجموعه آموزش  | میانگین سرعت باد |
| ۰/۵۶  | ۲/۶۰   | ۳/۵۹   | ۱/۱۶   | مجموعه آزمایش |                  |

جدول شماره ۲- معیارهای ارزیابی برای پیش‌بینی داده‌های سری زمانی تعداد موارد بروسلوز

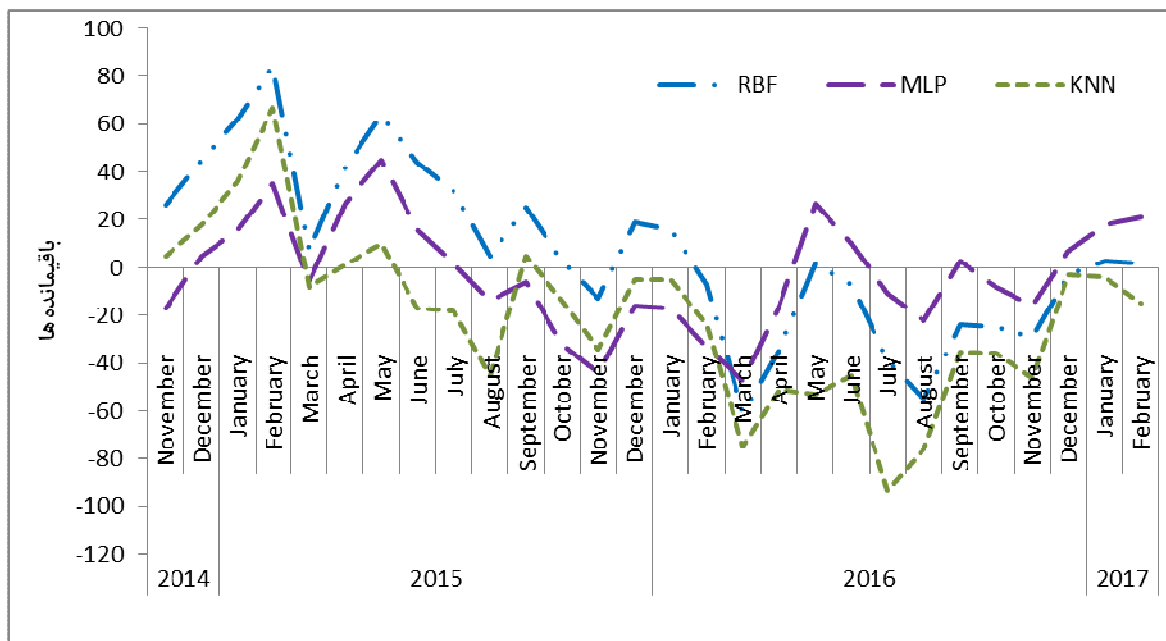
| ICC  | R2   | MRAE | MAE   | RMSE  | مدل                                   |
|------|------|------|-------|-------|---------------------------------------|
| ۰/۳۹ | ۰/۳۴ | ۰/۳۳ | ۲۹/۹۲ | ۳۶/۹۵ | شبکه عصبی مصنوعی مبتنی بر تابع رادیان |
| ۰/۷۵ | ۰/۶۱ | ۰/۲۵ | ۲۰/۶۵ | ۲۳/۷۹ | شبکه عصبی مصنوعی چند لایه پرسپترون    |
| ۰/۵۱ | ۰/۳۷ | ۰/۴۲ | ۳۲/۴۶ | ۴۰/۸۱ | K نزدیکترین همسایگی                   |



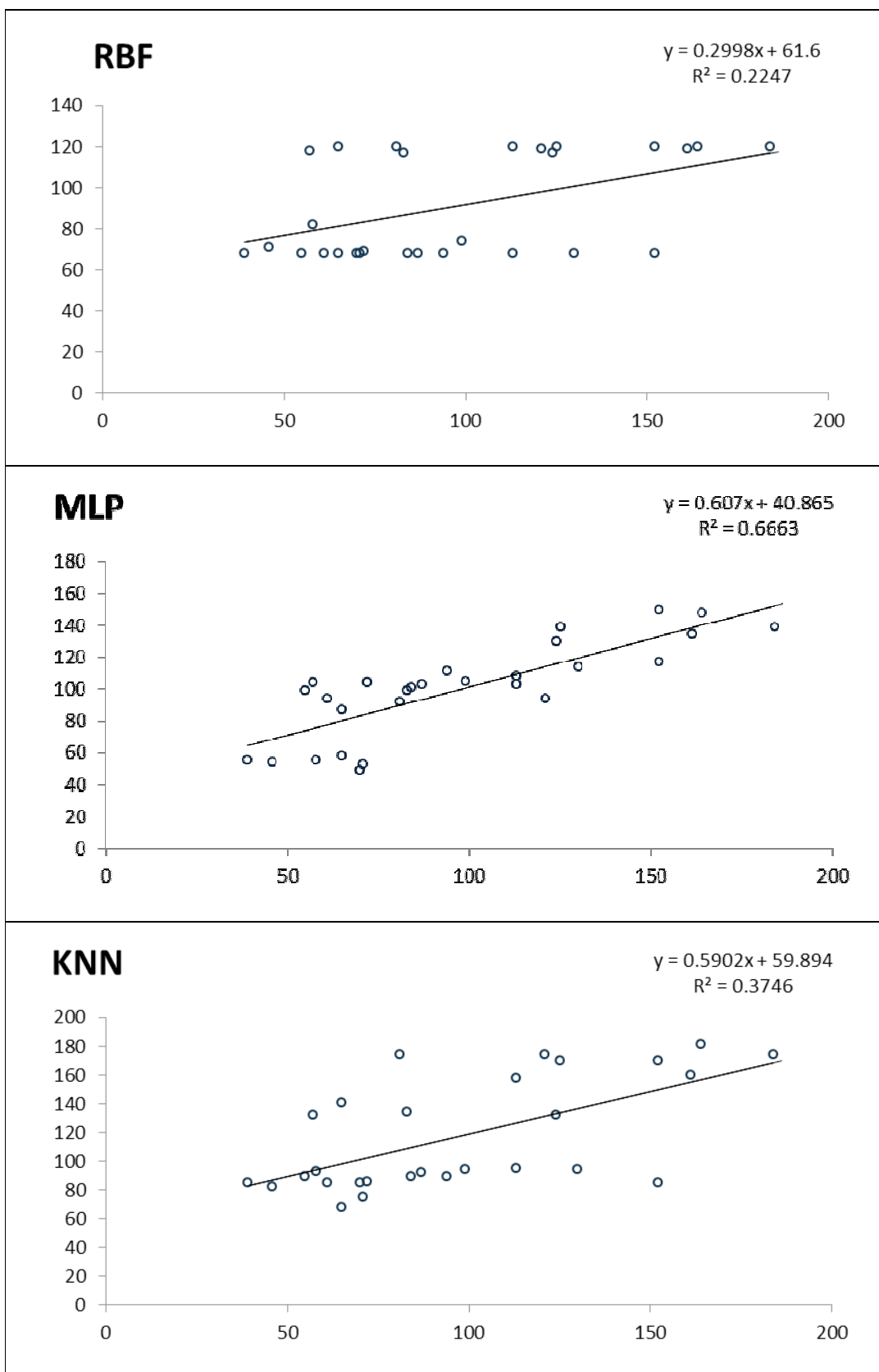
شکل شماره ۱- نمودار روند ماهانه تعداد موارد بروسلوز طی سال‌های ۱۳۸۳-۹۵



شکل شماره ۲- مقادیر پیش‌بینی شده با استفاده از RBF، MLP، KNN همراه با مقادیر مشاهده شده تعداد موارد ماهانه بروسلوز در مجموعه آزمایش



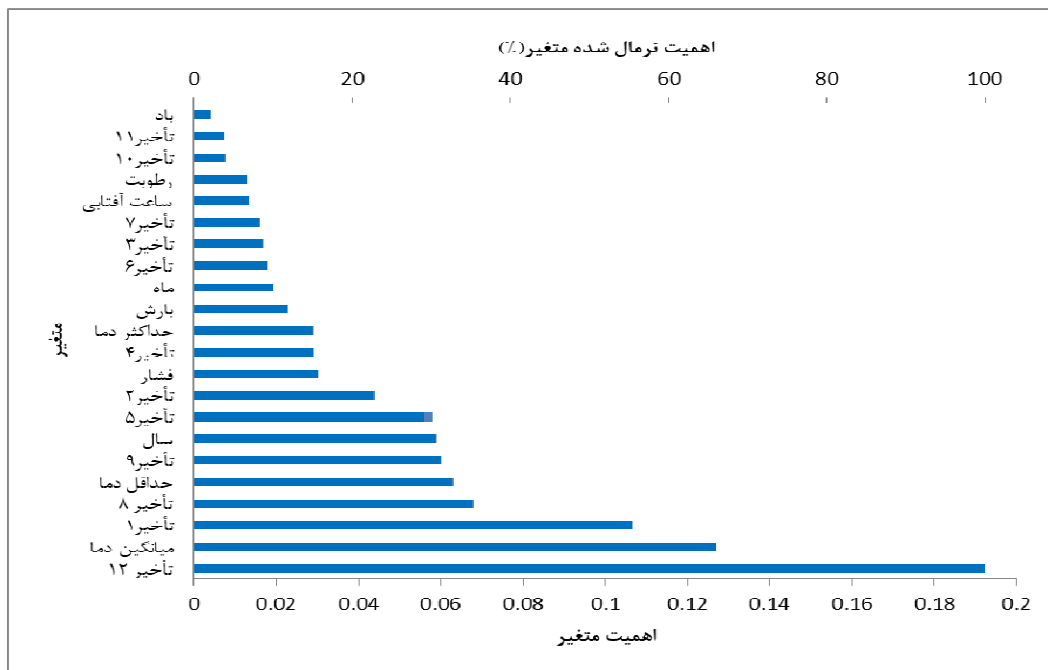
شکل شماره ۳- باقی‌مانده‌های حاصل از پیش‌بینی با استفاده از RBF، MLP، KNN



شکل شماره ۴- نمودار پراکنش مقادیر پیش‌بینی شده در مقابل مقادیر مشاهده شده تعداد موارد بروسلوز

Downloaded from irje.tums.ac.ir at 12:39 IRST on Monday December 10th 2018





شکل شماره ۵ - نمودار اهمیت متغیرها با استفاده از روش پرسپترون چندلایه

**بحث**

از روش MLP بود (۳۰). برآوردهای به دست آمده از روش KNN نیز تفاوت‌های زیادی از مقدارهای مشاهده شده داشتند. با این حال هیچ‌یک از روش‌ها مقادیر برازش شده منفی، که در واقعیت نیز رخداد آن غیرممکن است، برآورد نکردند. بر اساس یافته‌های معیارهای نیکویی برازش (MARE, MAE, MRSE, R<sup>2</sup>, ICC) و نمودارهای ترسیم شده مقادیر پیش‌بینی شده، این مطالعه به وضوح نشان داد که روش شبکه عصبی مصنوعی MLP در پیش‌بینی و نشان دادن روند دوره‌ای و غیردوره‌ای داده‌های سری زمانی ماهانه بروسلوز عملکرد بهتری نسبت به دو روش RBF و KNN داشت. یافته‌های این مطالعه نشان داد که روش MLP می‌تواند به‌طور کارایی در پیش‌بینی‌های ماهانه بروسلوز استفاده شود. شبکه‌های پرسپترون چند لایه از جمله مهم‌ترین ساختارهای شبکه‌های عصبی مصنوعی است که به‌طور معمول متشکل از چندین لایه ورودی هستند و سیگنال ورودی در خلال شبکه به‌صورت لایه به لایه منتشر می‌شود و بنابراین با توجه به دارا بودن ساختار پیچیده در پیش‌بینی متغیر خروجی از قابلیت تعمیم بهتری برخوردار است. این کار از طریق شناسایی تغییرات زمانی پیچیده درون داده‌های سری زمانی صورت می‌گیرد (۳۳). به‌تازگی مطالعه‌هایی در زمینه مقایسه عملکرد مدل‌های یادگیری ماشین در زمینه پیش‌بینی داده‌های سلامت در کشورهای مختلف انجام گرفته است. از جمله این مطالعه‌ها می‌توان به مطالعه ژانگ و همکاران (۱۰) اشاره نمود. در این

مدل‌های آماری با پیش‌بینی دقیق ابزار سودمندی در پیش‌بینی بروز بیماری در آینده هستند. روش‌های مبتنی بر تئوری یادگیری ماشین در سامانه‌های مراقبت بهداشتی بیماری‌های عفونی مختلف به‌طور مؤثر کارا بوده‌اند (۱۰، ۲۰، ۲۴، ۳۱). در این مطالعه، عملکرد روش‌های شبکه عصبی مصنوعی RBF و MLP و روش KNN در مدل‌سازی و پیش‌بینی تعداد موارد ماهانه بروسلوز بر اساس پارامترهای اقلیمی ارزیابی و مقایسه شد. به منظور افزایش عملکرد مدل‌ها، تأخیرهای ۱-۱۲ به‌عنوان متغیرهای پیش‌گو وارد مدل شد. ارزیابی عملکرد روش‌ها با استفاده از محاسبه معیارهای مختلف در مجموعه آزمون انجام شد. یافته‌ها نشان داد که روش MLP به کار رفته در این مطالعه به‌طور کارایی روند بیماری را مدل‌سازی کرده و در مقایسه با دو روش دیگر روند نزدیک‌تری به روند مشاهده شده داشت. روند پیش‌بینی شده با استفاده از روش شبکه عصبی RBF از روند مشاهده شده تفاوت زیادی داشت. ارزیابی بین روند مشاهده شده و پیش‌بینی شده منجر به برنامه‌ریزی‌های گمراه‌کننده در مراقبت بهداشتی خواهد شد (۳۲). مطالعه‌های زیادی وجود دارد که به‌منظور مقایسه عملکرد دو روش RBF و MLP انجام شده است. یافته این مطالعه با برخی از این مطالعه‌ها (۳۳) هم‌خوانی دارد، اما با مطالعه‌ی بایرام و همکاران هم‌سو نیست و در مطالعه آن‌ها عملکرد روش RBF بهتر

به‌طور معنی‌داری این بیماری بیش‌تر خواهد بود. به نظر می‌رسد تراکم در مراتع و بارندگی و دمای مناسب هم‌چنین رطوبت بالا و خاک مرطوب باعث رخداد بیش‌تر این بیماری خواهد بود. به عبارتی می‌توان گفت میانگین دمای استان که حدود ۱۲ درجه سانتی‌گراد است با تأخیر زمانی یک سال می‌تواند بیش‌ترین تأثیر در بین پارامترهای آب و هوایی در رخداد بروسلوز را داشته باشد. یعنی این باکتری در دمای ۱۲ درجه سانتی‌گرادی همدان به مدت یک سال در محیط باقی می‌ماند. آن‌چه مسلم است این که این باکتری‌ها در دمای حدی بالا و پایین عمر کم‌تری دارند. یعنی تابستان‌های داغ و زمستان‌های سرد باعث رخداد کم‌تر این بیماری می‌شود؛ در حالی که اعتدال آب و هوایی این دو فصل در منطقه همدان این بیماری را تشدید می‌کند. بر اساس یافته‌های به دست آمده پارامتر باد در سرعت‌های بالا این بیماری را کاهش می‌دهد و علت آن این است که این باکتری در هوا عمر کم‌تری دارد. با افزایش فشار تعداد افراد مبتلا بیش‌تر خواهد شد چرا که فشار بالا نشان از پایداری هوا دارد و به نظر می‌رسد در هوای نسبتاً پایدار و خاک مرطوب و دمای مناسب این بیماری تشدید می‌شود.

با ورود متغیرهایی مانند شاخص‌های شرایط اقتصادی و بهداشت حرفه‌ای و مقدار مصرف محصولات لبنی غیر پاستوریزه در مدل‌ها، ممکن است عملکرد برآزش‌ها بهبود یابد. از محدودیت‌های دیگر این مطالعه این بود که تنها از دو روش یادگیری ماشین استفاده شد و با هیچ مدل سری زمانی کلاسیک مقایسه انجام نشده است.

### نتیجه‌گیری

این مطالعه نشان داد که مدل شبکه عصبی مصنوعی MLP می‌تواند برای تشخیص رفتار تغییرات تعداد موارد بروسلوز انسانی در طول زمان و بر اساس تغییرات پارامترهای اقلیمی استفاده شود. بیش‌تر پارامترهای اقلیمی در رخداد این بیماری مؤثر بودند و مؤثرترین عامل در رخداد این بیماری دما بود. مطالعه‌های بیش‌تر در راستای کاربرد مدل‌های سری زمانی در عمل و شناسایی بهترین مدل به منظور کنترل و پیشگیری این بیماری کمک‌کننده خواهد بود.

### تشکر و قدردانی

نویسندگان بر خود لازم می‌دانند از حوزه‌ی بیماری‌های واگیر معاونت بهداشتی دانشگاه علوم پزشکی همدان به ویژه آقای

مطالعه روش‌های کلاسیک آریمما و هموارسازی نمایی با ماشین بردار پشتیبان مقایسه شد و ماشین بردار پشتیبان عملکرد بهتری را نشان داد. هم‌چنین گوان و همکاران در مطالعه‌ای به مقایسه عملکرد شبکه‌های عصبی با مدل‌های کلاسیک آماری برای پیش‌بینی بروز هپاتیت پرداختند و نشان دادند که روش شبکه‌های عصبی عملکرد بسیار بهتری نسبت به مدل‌های کلاسیک دارد (۳۴). اولیویرا و همکاران نیز در سال ۲۰۱۷ میلادی به مقایسه چند روش داده‌کاوی از جمله k- نزدیک‌ترین همسایگی و شبکه‌های عصبی پرسپترون چندلایه پرداختند. یافته‌های مطالعه آن‌ها نشان داد که از میان روش‌های مورد استفاده عملکرد روش پرسپترون بهتر از بقیه بود (۳۵). بنا به اطلاعات نگارندگان، این مطالعه نخستین مطالعه در ایران در زمینه کاربرد و مقایسه عملکرد روش‌های یادگیری ماشین در تحلیل داده‌های بروسلوز برای پیش‌بینی این بیماری است. پیشنهاد می‌شود مطالعه‌های آینده به مقایسه عملکرد سایر روش‌های داده‌کاوی در زمینه این بیماری در سایر استان‌های کشور بپردازند.

یکی دیگر از اهداف این پژوهش شناسایی عوامل اقلیمی و آب و هوایی مؤثر بروسلوز بود. زیرا با توجه به این‌که عامل بروسلوز باکتریایی است، شناسایی شرایط محیطی مناسب برای رشد این باکتری مانند شرایط دمایی، رطوبت، بارش، باد می‌تواند در مدیریت و کنترل این بیمار کمک شایانی نماید. بنابراین علاوه از متغیرهای تاخیرهای ۱-۱۲ ماه، در این پژوهش از داده‌های اقلیمی مانند دما، بارش، حداقل، حداکثر دما، رطوبت، باد، ساعت‌های آفتابی و فشار هوا استفاده شد و اثرات آن‌ها بر این بیماری با استفاده از روش‌های یاد شده مورد بررسی قرار گرفت. بر اساس یافته‌های به دست آمده از مدل پرسپترون چندلایه مشخص شد که عامل دما با وقوع بروسلوز ارتباط مستقیم دارد و نسبت به سایر پارامترهای اقلیمی مؤثرترین عامل در رخداد این بیماری است. در استان همدان که از نظر اقلیمی در منطقه نیمه خشک سرد واقع شده است و جزء مناطق کوهستانی کشور است، دامداری از رونق بالایی برخوردار است و به نظر می‌رسد شرایط اقلیمی همدان به دو راه مستقیم و غیر مستقیم نقش مؤثری در رخداد این بیماری دارد؛ چرا که در مواقعی که بارندگی و دما از شرایط مساعدی برخوردار است مراتع استان از کیفیت بالایی برخوردار خواهند بود که این خود رونق دامداری و افزایش تعداد زاد و ولد را منجر می‌شود. از طرفی بر اساس مطالعه آماری بین بارندگی و دما به طور مستقیم ارتباط معنی‌داری در رخداد بروسلوز وجود دارد، یعنی در سال‌هایی که بارندگی مناسب است

حمایت از مطالعه سپاس‌گزاری نمایند.

مهندس مسعود شجاعیان برای همکاری در استخراج داده‌ها و معاونت تحقیقات دانشگاه علوم پزشکی همدان به لحاظ تصویب و

## منابع

- Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC bioinformatics*. 2014; 15: 276.
- Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control: John Wiley & Sons; 2015. 5th edition. 4-17
- Moosazadeh M, Nikaeen R, Abedi G, Kheradmand M, Safiri S. Epidemiological and Clinical Features of People with Malta Fever in Iran: A Systematic Review and Meta-Analysis. *Osong Public Health and Research Perspectives*. 2016; 7: 157-67.
- Mirnejad R, Jazi FM, Mostafaei S, Sedighi M. Epidemiology of brucellosis in Iran: A comprehensive systematic review and meta-analysis study. *Microbial Pathogenesis*. 2017; 109: 239-47.
- Dean AS, Crump L, Greter H, Schelling E, Zinsstag J. Global Burden of Human Brucellosis: A Systematic Review of Disease Frequency. *PLOS Neglected Tropical Diseases*. 2012; 6: e1865.
- Eni P, Keramat F, Hasanzadeh Hoseinabadi M. Epidemiologic, clinical and laboratory findings of patients with brucellosis in Hamadan, west of Iran. *Journal of Research in Health Sciences*. 2012; 12: 105-8.
- Agasthya AS, Isloor S, Krishnamsetty P. Seroprevalence study of human brucellosis by conventional tests and indigenous indirect enzyme-linked immunosorbent assay. *The Scientific World Journal*. 2012; 2012: 104239.
- Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. *The Lancet Infectious Diseases*. 2006; 6: 91-9.
- Seyed Mohammad A, Mohammad Esmail M. A review of epidemiology, diagnosis and management of brucellosis for general physicians working in the Iranian health network. *Jundishapur Journal of Microbiology*. 2012; 384-7.
- Zhang X, Zhang T, Young AA, Li X. Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS One*. 2014; 9: e88075.
- González-Parra G, Arenas AJ, Jódar L. Piecewise finite series solutions of seasonal diseases models using multistage Adomian method. *Communications in Nonlinear Science and Numerical Simulation*. 2009; 14: 3967-77.
- Spaeder M, Fackler J. A multi-tiered time-series modelling approach to forecasting respiratory syncytial virus incidence at the local level. *Epidemiology & Infection*. 2012; 140(4):602-7.
- Li Q, Guo N-N, Han Z-Y, Zhang Y-B, Qi S-X, Xu Y-G, et al. Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. *The American Journal of Tropical Medicine and Hygiene*. 2012; 87: 364-70.
- Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. *Asian Pacific Journal of Tropical Medicine*. 2012; 5: 249-52.
- Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infectious Diseases*. 2011; 11: 218.
- Rios M, Garcia J, Sanchez J, Perez D. A statistical analysis of the seasonality in pulmonary tuberculosis. *European Journal of Epidemiology*. 2000; 16: 483-8.
- Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *The American Journal of Tropical Medicine and Hygiene*. 2008; 79: 933-9.
- Yu L, Zhou L, Tan L, Jiang H, Wang Y, Wei S, et al. Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. *PLoS one*. 2014; 9: e98241.
- Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X. Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PLoS one*. 2013; 8: e63116.
- Ture M, Kurt I. Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Systems with Applications*. 2006; 31: 41-6.
- Trtica-Majnaric L, Zekic-Susac M, Sarlija N, Vitale B. Prediction of influenza vaccination outcome by neural networks and logistic regression. *Journal of Biomedical Informatics*. 2010; 43: 774-81.
- Baboo SS, Shereef IK. An efficient weather forecasting system using artificial neural network. *International Journal of Environmental Science and Development*. 2010; 1: 321.
- Amid S, Mesri Gundoshmian T, Shahgoli G. Comparison of MLP and RBF neural networks performance for estimation of broiler output energy. *Iranian Journal of Biosystems Engineering*. 47: 319-28.
- Chakraborty P, Khadivi P, Lewis B, Mahendiran A, Chen J, Butler P, et al., editors. Forecasting a moving target: Ensemble models for IIL case count predictions. *Proceedings of the 2014 SIAM international conference on data mining*; 2014: SIAM.
- Zeinali M, Shirzadi M, Sharifian J. National guideline for brucellosis control. Tehran: Ministry of Health and Medical Education. 2009: 10-7.
- Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning*: Springer series in statistics New York; 2001.
- Tominola M, Tynkkynen M, Lemmetty J, Harstel P, Sikanen L. Estimating the characteristics of a marked stand using k-nearest-neighbour regression. *Journal of Forest Engineering*. 1999; 10: 75-81.
- Yoon H, Jun S-C, Hyun Y, Bae G-O, Lee K-K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology*. 2011; 396: 128-38.
- Tapak L, Mahjub H, Hamidi O, Poorolajal J. Real-data comparison of data mining methods in prediction of diabetes in Iran. *Healthcare informatics research*. 2013; 19: 177-85.
- Bayram S, Ocal ME, Laptali Oral E, Atis CD. Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey. *Journal of Civil Engineering and Management*. 2016; 22: 480-90.
- Song Q, Zheng Y-J, Xue Y, Sheng W-G, Zhao M-R. An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. *Neurocomputing*. 2017; 226: 16-22.
- Zhang X, Zhang T, Pei J, Liu Y, Li X, Medrano-Gracia P. Time series modelling of syphilis incidence in China from 2005 to 2012. *PLoS One*. 2016; 11: e0149401.
- Memarian H, Balasundram SK. Comparison between multi-layer perceptron and radial basis function networks for

- sediment load estimation in a tropical watershed. *Journal of Water Resource and Protection*. 2012; 4: 870.
34. Guan P, Huang D-S, Zhou B-S. Forecasting model for the incidence of hepatitis A based on artificial neural network. *World journal of gastroenterology: WJG*. 2004; 10: 3579.
35. Oliveira A, Faria BgMn, Gaio AR, Reis LsP. Data Mining in HIV-AIDS Surveillance System. *Journal of medical systems*. 2017; 41: 51

Iranian Journal of Epidemiology Summer 2018; 14 (2): 1-8.

Original Article

# Predicting the Frequency of Human Brucellosis using Climatic Indices by Three Data Mining Techniques of Radial Basis Function, Multilayer Perceptron and Nearest Neighbor: A Comparative Study

Tapak L<sup>1</sup>, Shirmohammadi-Khorram N<sup>2</sup>, Hamidi O<sup>3</sup>, Maryanaji Z<sup>4</sup>

1- PhD of Biostatistics, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

2- Msc of Biostatistics, Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

3- Department of Science, Hamedan University of Technology, Hamedan, Iran

4- PhD of Climatology, Department of Geography, Sayyed Jamaledin Asadabadi University, Asadabad, Iran

**Corresponding author:** Shirmohamadi N, n.shirmohamadi.kh@gmail.com

(Received 17 September 2017; Accepted 3 March 2018)

**Background and Objectives:** Identification of statistical models has a great impact on early and accurate detection of outbreaks of infectious diseases and timely warning in health surveillance. This study evaluated and compared the performance of the three data mining techniques in time series prediction of brucellosis.

**Methods:** In this time series, the data of the human brucellosis cases and climatology parameters of Hamadan, west of Iran, were analyzed on a monthly basis from 2004 (March/April) to 2017 (February/March). The data were split into two subsets of train (80%) and test (20%). Three techniques, i.e. radial basis function (RBF) and multilayer perceptron (MLP) artificial neural network methods as well as K Nearest neighbor (KNN), were used in both subsets. The root mean square errors (RMSE), mean absolute errors (MAE), mean absolute relative errors (MARE), determination coefficient ( $R^2$ ) and intra-class correlation coefficient (ICC) were used for performance comparison.

**Results:** Results indicated that RMSE (23.79), MAE (20.65) and MARE (0.25) for MLP were smaller compared to the values of the other two models. The ICC (0.75) and  $R^2$  (0.61) values were also better for this model. Thus, the MLP model outperformed the other models in predicting the used data. The most important climatology variable was temperature.

**Conclusion:** MLP can be effectively applied to diagnose the behavior of brucellosis over time. Further research is necessary to detect the most suitable method for predicting the trend of this disease.

**Keywords:** Human Brucellosis, Artificial neural networks, Nearest neighbor, Prediction