مجله دانشکده پزشکی، دانشگاه علوم پزشکی تهران، تیر ۱۳۹۸، دوره ۷۷، شماره ٤، صفحه‌های ۲۱٦ تا ۲۲۱

مقاله اصیل

پیش‌بینی میزان بروز بیماری سل با استفاده از سری‌های زمانی مبتنی بر شبکه‌های عصبی در ایران

عاطفه صدیق‌نیا[1*]
شراره رستم نیاکان کلهری[1]
مهشید ناصحی[2]
احمد علی حنفی بجد[3]

۱- گروه انفورماتیک پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران؛ گروه مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران.
۲- گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، مرکز مدیریت بیماری‌های واگیر وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران.
۳- گروه حشره‌شناسی پزشکی و مبارزه با ناقلین، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران.

[*] نویسنده مسئول: تهران، خیابان انقلاب، خیابان قدس، فردانش، طبقه سوم ساختمان شماره ۱، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، گروه انفورماتیک پزشکی.
تلفن: ۸۸۹۸۲۸۸٦ -۰۲۱
E-mail: asedighnia@gmail.com

چکیده

دریافت: ۱۳۹۸/۰۲/۲٤ ویرایش: ۱۳۹۸/۰۲/۳۱ پذیرش: ۱۳۹۸/۰٤/۲٤ آنلاین: ۱۳۹۸/۰٤/۳۱

زمینه و هدف: یکی از بیماری‌های عفونی مهم با مرگ‌ومیر بالا در جهان، سل می‌باشد که هیچ کشوری از آن مصون نیست و امروزه به‌علل مختلف مانند بیماری‌های زمینه‌ای بروز آن بار دیگر در حال افزایش می‌باشد. براساس آخرین گزارش سازمان بهداشت جهانی از وضعیت سل در ایران، سل مقاوم به دارو (MDR-TB) و سل در افراد دارای ویروس نقص ایمنی انسانی (Human immunodeficiency virus, HIV) در کشور رو به افزایش است. پیش‌بینی بروز برای پیشگیری، مدیریت و کنترل بهتر این بیماری امری لازم می‌باشد. هدف این مطالعه ایجاد سیستم پیش‌بینی کننده میزان بروز سل می‌باشد.

روش بررسی: تحلیل گذشته‌نگری بر روی ۱۰٦۵۱ بیمار مسلول ثبت شده بین اول فروردین ۱۳۹۳ تا پایان اسفند ۱۳۹٤ در سیستم وزارت بهداشت، درمان و آموزش پزشکی ایران انجام گرفت. پارامترهای قابل استناد جداسازی شدند و به‌صورت مستقیم، ادغام و یا تولید شاخص جدید در نظر گرفته شدند.

یافته‌ها: ۲۳ متغیر مستقل وارد مطالعه شد که با سنجش همبستگی و محاسبه رگرسیون، ۱۲ متغیر با P≤۰/۰۱ در اسپیرمن و P≤۰/۰۵ در پیرسون مرتبط شناخته شد. بهترین نتایج R=۰/۹۳ و MSE=۱۰/۹٦ در داده‌های آموزش، صفر و صفر در داده‌های اعتبارسنجی و R=۰/۹۱ و MSE=۱۳/۲۳ در داده‌های تست و همچنین نمودار رگرسیون چشمگیری از شبکه ایجاد شده با الگوریتم‌های سری زمانی شبکه عصبی در متلب به‌دست آمد.

نتیجه‌گیری: نتایج مطالعه حاضر بیانگر این است که هوش مصنوعی برای استخراج دانش از داده‌های خام جمع‌آوری شده مربوط به بیماری سل عملکرد مناسبی دارد و می‌توان از آن برای پیش‌بینی موارد جدید این بیماری استفاده کرد.

کلمات کلیدی: بروز، ایران، شبکه عصبی، پژوهش‌های گذشته‌نگر، بیماری سل.

مقدمه

پایش و مراقبت هسته فعالیت‌های بهداشت عمومی برای کنترل و ریشه‌کن سازی سل است.[1] بروز در واقع موارد جدید مبتلا شده در هر سال می‌باشد که بررسی بروز سالانه بیماری‌ها در سراسر جهان برای پیش‌بینی مناسب هستند. از این پیش‌بینی‌ها می‌توان برای برنامه‌ریزی جهت جلوگیری از شیوع بیماری‌ها استفاده کرد، همچنین پیش‌بینی اپیدمی بیماری‌ها باعث عملکرد مناسب و به موقع اقدامات پیشگیری می‌شود.[2] به‌وسیله بررسی داده‌های بیماری سل، افزون‌بر پیشگیری از مرگ ناشی از این بیماری، می‌توان مسایل دیگری مانند شکست سیستم سلامت و شبکه اجتماعی و دلایل افزایش موارد بیماران و تعداد آن‌ها را از نظر اجتماعی- جمعیتی، بالینی و در معرض خطر بودن را نیز یافت.[3و4] مدل‌های مبتنی بر سری‌های زمانی برای پیش‌بینی موارد و شیوع‌های قریب‌الوقوع بیماری‌های عفونی

دارای الگوهای فصلی و افزایش دقت این پیش‌بینی‌ها مفید هستند.[۶و۵] استفاده همزمان از روش‌های هوش مصنوعی و سیستم‌های اطلاعاتی منجر به ایجاد سیستم‌های هوشمند می‌شود که برای حل مسائل پیچیده کاربردی هستند.[۸و۷] شبکه عصبی مصنوعی که فعالیت آن براساس شبکه‌های عصبی بیولوژیکی است، با توانایی چشمگیر در یافتن روابط غیر خطی مابین متغیرهای مستقل با یکدیگر و با متغیر وابسته به وسیله‌ی وزن‌دهی به ورودی‌ها، عملکرد مناسبی در دسته‌بندی و پیش‌بینی دارند. شبکه‌های عصبی از داده‌های گذشته برای آموزش استفاده می‌کنند و می‌توانند با توجه به ورودی‌های جدید مقدار خروجی را تعیین کنند.[۱۰و۹]

بدیهی است با توجه به روند جهانی معکوس بیماری سل و عوامل تاثیرگذار مختلف بر این بیماری، جهت کنترل بهتر سل و تخصیص موثر منابع موجود، بررسی تغییرات زمانی بروز بیماری و پیش‌بینی روندهای آن در آینده امری ضروری می‌باشد.[۱۱] مطالعه کنونی با هدف پیش‌بینی میزان بروز بیماری سل انجام گردید.

روش بررسی

در این مطالعه به‌صورت گذشته‌نگر داده‌های دریافتی مربوط به بیماران مسلول ثبت شده در سیستم ثبت بیماری سل کشور ایران که طی سال‌های ۱۳۹۳ تا ۱۳۹٤ تشخیص داده شده‌اند، در نظر گرفته شد. داده‌های مربوط به ۱۰٦۵۱ مورد مبتلا به بیماری سل با در نظر گرفتن تاریخ تشخیص ثبت شده در سامانه، براساس فصول سال به ۸ دسته تقسیم شدند. متغیرهای قابل استناد براساس میزان داده‌های ثبتی، جداسازی شد و ۳ متغیر محیطی تاثیرگذار استخراج شده از متون علمی به پارامترهای پیشین افزوده شد. پارامترهای محیطی شامل میانگین دمای فصلی، میانگین سرعت باد در هر فصل سال و وضعیت اقلیمی مربوط به شهر محل قرارگیری هر کدام از دانشگاه‌های علوم پزشکی بود. در ابتدا ضریب همبستگی هر متغیر مستقل به‌صورت مجزا با متغیر وابسته سنجیده شد سپس با توجه به ماهیت متغیر وابسته، برای انتخاب متغیرهای تاثیرگذار از رگرسیون خطی با تعداد متغیرها و روش‌های ورود داده SPSS statistical software, version 20 (IBM, Armonk, NY, USA) استفاده شد و پارامترهای حاصل از مدل با بهترین R جهت ورود به شبکه انتخاب شدند. مقادیر مربوط به پارامترهای انتخابی برای آموزش به سری‌های زمانی شبکه عصبی در نرم‌افزار MATLAB® software, version R2014a (Mathworks Inc., Natick, MA, USA) داده شد. برای دسترسی به شبکه‌ای با بهترین کارکرد، شبکه‌هایی با الگوریتم‌ها و ساختارهای مختلف ایجاد شد که عملکرد آن‌ها در مجموعه داده‌های آموزش (۷۵٪)، اعتبارسنجی (۱۵٪) و داده‌های تست (۱۵٪) براساس مقادیر R و MSE (Mean squared error) و نمودار رگرسیون (Logistic regression model) مورد ارزیابی قرار گرفت.

یافته‌ها

در ابتدا ۲۳ متغیر که مجموع پارامترهای قابل استناد استخراجی از فایل دریافتی از وزارت بهداشت، درمان و آموزش پزشکی کشور و داده‌های محیطی که در منابع علمی پیشین به آن‌ها اشاره شده بود، انتخاب شدند. گفتنی است که همه متغیرها به جز نتیجه اسمیر و شاخص توده بدنی به‌صورت مستقیم وارد مطالعه شد. شاخص توده بدنی (Body mass index, BMI) حاصل انجام محاسباتی بر روی متغیرهای پیشین قد و وزن برای تولید متغیر جدید بود و نتیجه اسمیر پیش از درمان آزمایشگاه‌های محلی و رفرانس با یکدیگر ادغام شدند که در جدول ۱ به آن‌ها اشاره شده است. پارامترهای ورودی شامل: نام دانشگاه، وضعیت تاهل، جنسیت، سابقه زندان، سن، ملیت، تحصیلات، شغل، محل سکونت، نوع بیماری، درجه اسمیر، نتیجه رادیوگرافی، نتیجه درمان، ابتلا به ایدز، سابقه مواجهه با فرد مسلول و زمان مواجهه، بستری مرتبط و مدت آن، شاخص توده بدنی، فصل، سرعت باد، میانگین دما و وضعیت اقلیمی می‌باشد.

داده‌های کمی موجود در این مطالعه از جمله متغیر وابسته دارای توزیع غیرنرمال بودند، به‌همین جهت ضریب همبستگی اسپیرمن با P≤۰/۰۵ در نظر گرفته شد، سپس مدل رگرسیون خطی مربوط به آن‌ها به دو روش گام به گام و ورودی ایجاد شد که بهترین نتایج حاصل از مدلی با روش گام به گام با R=۰/۵۸۰ بود که ۱۲ متغیر را مرتبط تشخیص داد که به‌عنوان ورودی شبکه عصبی در نظر گرفته شدند. شبکه عصبی مبتنی بر سری‌های زمانی که در مطالعه حاضر ایجاد شده است، حاصل از شبکه واپسگرای خودکار غیرخطی با ورودی‌های خارجی بود. در این شبکه مقدار سری جدید براساس

جدول ۱: متغیرهای منتخب جهت ورود به مطالعه

نام متغیر	وضعیت ورود به مطالعه
نام دانشگاه	ورود مستقیم
وضعیت تاهل	ورود مستقیم
جنسیت	ورود مستقیم
سابقه زندان	ورود مستقیم
سن	ورود مستقیم
قد	متغیر جدید شاخص توده بدنی
وزن	
ملیت	ورود مستقیم
میزان تحصیلات	ورود مستقیم
شغل	ورود مستقیم
محل سکونت	ورود مستقیم
تاریخ تشخیص	معیار دسته بندی
نوع بیماری	ورود مستقیم
مورد بیماری	ورود مستقیم
نتیجه رادیوگرافی قفسه سینه	ورود مستقیم
آلودگی به HIV	ورود مستقیم
نتیجه اسمیر آزمایشگاه رفرنس	
پیش از درمان	ترکیب جهت ایجاد متغیر نتیجه اسمیر
نتیجه اسمیر آزمایشگاه محلی	پیش از درمان
پیش از درمان	
نتیجه درمان	ورود مستقیم
سابقه بستری مرتبط با بیماری سل	ورود مستقیم
مدت بستری	ورود مستقیم
سابقه تماس با بیمار مبتلا به سل	ورود مستقیم
زمان سابقه تماس	ورود مستقیم

جدول ۲: متغیرهای تاثیرگذار گزارش شده بر بروز بیماری سل در مطالعه حاضر و مطالعات پیشین

متغیرهای موثر گزارش شده	تعداد منابع گزارش کننده	حضور در مطالعه حاضر
وضعیت تاهل	۵	*
جنسیت	٤	*
سابقه زندان	۲	*
سن	۵	*
ملیت	۷	*
تحصیلات	۲	*
شغل	۲	*
محل سکونت	۲	*
نوع بیماری	٤	*
نتیجه اسمیر	۲	*
رادیوگرافی	۱	*
نتیجه درمان	٤	*
مواجهه با مسلول و زمان آن	۳	*
بستری مرتبط با سل و مدت آن	۱	*
شاخص توده بدنی	۱	*
فصل	۷	*
سرعت باد	۲	*
دما	۲	*
دسته‌بندی اقلیمی	۳	*
ایدز	۸	*
دیابت	۳	–
مصرف الکل	۳	–
مواد دخانی	٤	–

مقادیر پیشین همان سری و سری‌های دیگر پیش‌بینی می‌شود. متغیرهای حاصل از رگرسیون خطی در ماتریسی به‌عنوان ورودی وارد شدند که دارای ابعاد ۱۰٦۵۱ سطر و ۱۲ ستون بود. مقادیر بروز سل نیز در ماتریس خروجی قرار گرفتند که ۱۰٦۵۱ سطر در ۱ ستون بود. ۷۰٪ داده‌ها برای آموزش و ۱۵٪ برای اعتبارسنجی و ۱۵٪ برای تست در نظر گرفته شدند. برای ایجاد شبکه به روش آزمون و خطا تعداد نورون‌های لایه مخفی را به‌صورت پلکانی افزایش داده و برای تاخیرهای ۱ و ۲ الگوهای بالا را با الگوریتم‌های مختلف به‌صورت مجزا پیاده‌سازی کردیم که در نهایت ٦٦ شبکه آموزش دیده با داده‌های مطالعه کنونی تولید شد. بهترین نتایج از الگوریتم رگولاریزیشن بیزین (Bayesian regularization) با ۱۰ نرون در لایه مخفی و تاخیر ۲ با R=۰/۹۱ و MSE=۱۳/۲۳ در داده‌های تست، R=۰/۰ و MSE=۰/۰ در داده‌های اعتبارسنجی و R=۹/۳۳ و MSE=۱۰/۹٦ در مجموعه آموزش به‌دست آمد. همچنین مدت زمان صرف شده برای آموزش این شبکه ۲ دقیقه و ٦ ثانیه به طول انجامید. نمودار رگرسیون مربوط به کل داده‌ها در نمودار ۱ نشان داده شده است.

بحث

در این مطالعه در ابتدا ۲۳ متغیر به‌عنوان عوامل موثر بر بروز بیماری سل در نظر گرفته شدند. مطالعات پیشین تاثیرگذار بودن همه این متغیرها را بر بروز این بیماری گزارش کرده‌اند که خلاصه آن‌ها در جدول ۲ نشان داده شده است. با توجه به مطالعه پژوهش‌های پیشین انجام شده و جستجو در سایت‌های علمی بین‌المللی، شامل PubMed و Scopus که مربوط به مقالات علمی است و OPPO که مربوط به اختراع است، باید در نظر داشت که تاکنون سیستمی برای پیش‌بینی بروز بیماری سل بر پایه سری‌های زمانی شبکه عصبی در دنیا طراحی نشده است. به همین دلیل مطالعه حاضر شروعی برای مدل‌سازی چنین سامانه‌هایی می‌باشد. یافته‌های مطالعه‌ای که برای پیش‌بینی بروز سل در منطقه‌ای از چین ٤٦٥٩٦٠ نمونه را در نظر گرفتند، نشان می‌دهد که خطای پیش‌بینی در ترکیب شبکه‌های عصبی با سری زمانی کمتر از مقدار این شاخص در مدل‌های سری‌های زمانی به تنهایی می‌باشد.[۱۲]

مطالعه دیگری با موضوع "پیش‌بینی سل فعال ریوی با استفاده از شبکه عصبی مصنوعی" توسط El-Solh و همکاران با شبکه عصبی رگرسیون عمومی انجام شد که به ۰/۹٤ دقت دست یافتند.[۱۳] Seixas و همکاران با ۱۳۷ بیمار شبکه عصبی پرسپترون چند لایه پیشخور برای تشخیص سل ریوی ایجاد کردند که به ۹۰٪ دقت در داده‌های آزمون رسیدند.[۱٤] Elveren و همکاران نیز برای تشخیص بیماری سل از الگوریتم ژنتیک شبکه عصبی استفاده کردند که دقت مطالعه ۹٤٪ گزارش شده است و زمان مورد نیاز برای آموزش زیاد بوده است. این مطالعه بیان دارد که استفاده از ترکیب چند روش نتایج بهتری خواهد داشت.[۱۵] در مطالعه حاضر، خروجی شبکه مقدار عددی می‌باشد. حایز اهمیت است که با در اختیار داشتن اعداد دقیق، تصمیم‌گیرنده با اطمینان بهتر می‌تواند تصمیم‌گیری نماید تا این که فقط تعدادی گروه از پیش تعیین‌شده به او گزارش شود. چنین نحوه پیش‌بینی نیز تاکنون برای بیماری سل انجام نشده است. در سال ۲۰۱۷ مطالعه‌ای برای ارزیابی کلسترول و تری‌گلیسیرید در افراد دارای اضافه وزن به‌وسیله رگرسیون خطی چندگانه و مدل هوش مصنوعی انجام گرفت. در شبکه عصبی پس انتشار خطا بهترین شبکه در پیش‌بینی تری‌گلیسیرید در تکرار ۵۳ با R=۰/۹۹۹۷ به‌دست آمد در حالی‌که برای کلسترول نتیجه چشمگیری پس از تکرار ۱۰۰۰ با R=۰/۹۹۲۲ به‌دست آمد. این در حالی است که در مطالعه حاضر بهترین نتایج در تکرار ٤۱۹ به‌دست آمد.[۱٦] در پژوهش و همکاران، La Delfa مقدار قدرت بازو با دو روش رگرسیون چند متغیره و شبکه عصبی پیش‌بینی شد که با ٤٥٦ داده آموزش، در شبکه عصبی r=٪۹۰/۲ و اصل مجموع مربعات انحراف‌ها برابر بود با ۹/۳ در حالی‌که این مقادیر در رگرسیون برای داده‌های آموزش ٦٦/۵٪ و ۱۷/۲ بود.[۱۷] Wang و همکاران برای پیش‌بینی مقدار گلوکز خون در بیماران دارای اضافه وزن از روشی مشابه روش مطالعه حاضر استفاده کرده‌اند. آن‌ها برای انتخاب متغیرهای موثر از رگرسیون خطی چندگانه استفاده کردند و سپس برای پیش‌بینی، مقادیر داده‌ای مربوط به ٦ متغیر ۳٤٦ بیمار را برای آموزش به شبکه عصبی پس انتشار خطا دادند که پس از ۱۰۰۰ بار تکرار به دقت ۹۹/۸۷٪ رسید.[۱۸]

در مطالعه‌ای که در سال ۲۰۱٤ به‌صورت مشترک بین چین و روسیه با استفاده از داده‌های سری زمانی ایستگاه کنترل دریای زرد انجام شد. در فرآیندهای زیست دریایی عوامل زیادی از جمله بیولوژیکی و انسانی و هواشناسی و هیدرولوژیکی تاثیرگذار هستند که مدل‌های معمولی به سختی می‌توانند آن را تحلیل کنند، به همین دلیل مدل یکپارچه شده سری‌های زمانی و شبکه عصبی غیرخطی برای حل این مسئله پیشنهاد شده است. نتایج این مطالعه حاکی از آن است که مدل‌های پیش‌بینی با شبکه عصبی که از داده‌های سری زمانی استفاده کنند، برای ایجاد سیستم‌های هشدار سریع مفید و موثر خواهند بود که می‌تواند با هشدار سریع از پیامدهای ناگوار آینده جلوگیری کند. [۱۹] با توجه به مقایسه مطالعات پیشین با نتایج مطالعه حاضر مشخص می‌شود که سری زمانی شبکه عصبی روش مناسبی برای پیش‌بینی میزان بروز بیماری سل است که می‌تواند به‌عنوان پیش‌آگهی پیک‌های این بیماری و مناطق در معرض خطر بیشتر به شمار برود و به این وسیله بر مدیریت بیماری سل تاثیر بسزایی بگذارد. پیشنهاد می‌شود از سایر روش‌ها و الگوریتم‌های شبکه عصبی نیز برای پیش‌بینی روندهای این بیماری استفاده شود و مطالعات مقایسه‌ای بین آن‌ها انجام گیرد. با توجه به نتایج به‌دست آمده در مطالعه حاضر، به‌علت وجود روند فصلی در بیماری سل، سری زمانی روش مناسبی برای پیش‌بینی بروز این بیماری است. همچنین شبکه‌های عصبی مصنوعی برای استخراج دانش از داده‌های خام

مربوط به بیماری سل عملکرد قابل پذیرشی دارد به‌همین دلیل می‌توان از آن برای پیش‌بینی موارد جدید این بیماری استفاده کرد.

سپاسگزاری: این مطالعه بخشی از پایان‌نامه تحت عنوان "طراحی و ایجاد سامانه پیش‌بینی میزان بروز بیماری سل با استفاده از سری‌های زمانی مبتنی بر شبکه‌های عصبی و سیستم اطلاعات جغرافیایی در ایران" در مقطع کارشناسی ارشد انفورماتیک پزشکی دانشکده پیراپزشکی در سال ۹۶ با کد ۳۷/الف/۲۸۰/۳ می‌باشد که با حمایت دانشگاه علوم پزشکی تهران اجرا شده است.

References

1. Castro KG. Tuberculosis surveillance: data for decision-making. *Clin Infect Dis* 2007;44(10):1268-70.
2. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infect Dis* 2011;11:166.
3. Dye C, Bassili A, Bierrenbach AL, Broekmans JF, Chadha VK, Glaziou P, et al. Measuring tuberculosis burden, trends, and the impact of control programmes. *Lancet Infect Dis* 2008;8(4):233-43.
4. Ghosh S, Moonan PK, Cowan L, Grant J, Kammerer S, Navin TR. Tuberculosis genotyping information management system: enhancing tuberculosis surveillance in the United States. *Infect Genet Evol* 2012;12(4):782-8.
5. Bhatnagar S, Lal V, Gupta SD, Gupta OP. Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian J Public Health* 2012;56(4):281-5.
6. Kennedy CE, Aoki N, Mariscalco M, Turley JP. Using time series analysis to predict cardiac arrest in a PICU. *Pediatr Crit Care Med* 2015;16(9):e332-9.
7. Kunhimangalam R, Ovallath S, Joseph PK. A novel Fuzzy Expert System for the identification of severity of carpal tunnel syndrome. *Biomed Res Int* 2013;2013:846780.
8. Cao Y, Hu ZD, Liu XF, Deng AM, Hu CJ. An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. *Dis Markers* 2013;35(6):653-60.
9. Sarlak MA, Forati H, editors. Advanced Management Information Systems. Tehran, Iran: Payam Noor University; 2009. [Persian]
10. Sheikhtaheri A, Sadoughi F, Hashemi Dehaghi Z. Developing and using expert systems and neural networks in medicine: a review on benefits and challenges. *J Med Syst* 2014;38(9):110.
11. Biglarian A, Babaei Rochi Gh, Azmi R. Application of artificial neural network in determining the most important predictor of mortality within the hospital after open heart surgery and comparison with logistic regression model. *Modares J Med Sci* 2004;7(1):9-23. [Persian]
12. 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(5):e63116.
13. El-Solh AA, Hsiao CB, Goodnough S, Serghani J, Grant BJ. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest* 1999;116(4):968-73.
14. Seixas JM, Faria J, Souza Filho JB, Vieira AF, Kritski A, Trajman A. Artificial neural network models to support the diagnosis of pleural tuberculosis in adult patients. *Int J Tuberc Lung Dis* 2013;17(5):682-6.
15. Elveren E, Yumuşak N. Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. *J Med Syst* 2011;35(3):329-32.
16. Ma J, Yu J, Hao G, Wang D, Sun Y, Lu J, et al. Assessment of triglyceride and cholesterol in overweight people based on multiple linear regression and artificial intelligence model. *Lipids Health Dis* 2017;16(1):42.
17. La Delfa NJ, Potvin JR. Predicting manual arm strength: A direct comparison between artificial neural network and multiple regression approaches. *J Biomech* 2016;49(4):602-5.
18. Wang J, Wang F, Liu Y, Xu J, Lin H, Jia B, et al. Multiple linear regression and artificial neural network to predict blood glucose in overweight patients. *Exp Clin Endocrinol Diabetes* 2016;124(1):34-8.
19. Zhang Y, Wang J, Vorontsov AM, Hou G, Nikanorova MN, Wang H. Using a neural network approach and time series data from an international monitoring station in the Yellow Sea for modeling marine ecosystems. *Environ Monit Assess* 2014;186(1):515-24.

TUMJ Tehran University Medical Journal, July 2019; Vol. 77, No. 4: 216-221 Original Article

Tuberculosis incidence predicting system using time series neural network in Iran

Atefeh Sedighnia M.Sc.[1*]
Sharareh Rostam Niakan Kalhori Ph.D.[1]
Mahshid Nasehi Ph.D.[2]
Ahmad Ali Hanafi-Bojd Ph.D.[3]

1- Department of Medical Informatics, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran; Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran.
2- Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Infectious Diseases Management Center of Ministry of Health and Medical Education, Tehran, Iran.
3- Department of Medical Entomology, Tehran University of Medical Sciences, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.

[*]Corresponding author: Department of Medical Informatics, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran.
Tel: +98- 21- 88982886
E-mail: asedighnia@gmail.com

Abstract

Received: 14 May 2019 Revised: 21 May 2019 Accepted: 15 Jul. 2019 Available online: 22 Jul. 2019

Background: Tuberculosis (TB) is an important infectious disease with high mortality in the world. None of the countries stay safe from TB. Nowadays, different factors such as co-morbidities, increase TB incidence. World Health Organization (WHO) last report about Iran's TB status shows rising trend of multidrug-resistant tuberculosis (MDR-TB) and HIV/TB. More than 95% illness and death of TB cases are in developing countries. The most infections are in South East Asia and West Pacific that 56% of them are new cases in the world. The incidence is actually new cases of each year. Incidence prediction is affecting TB prevention, management and control. The purpose of this study was to designing and creating a system to predict TB incidence by time series artificial neural networks (ANN) in Iran.

Methods: This study was a retrospective analytic. 10651 TB cases that registered on Iran's Stop TB System from March 2014 to March 2016, were analyzed. Most of reliable data used directly, some of them merged together and create new indicators and two columns used to compute a new indicator. At first, effective variables were evaluating with correlation coefficient tests then extracting by linear regression on SPSS statistical software, version 20 (IBM, Armonk, NY, USA). We used different algorithms and number of neurons in hidden layer and delay in time series neural network. R, MSE (mean squared error) and regression graph were used for compare and select the best network. Incidence prediction neural network were designed on MATLAB® software, version R2014a (Mathworks Inc., Natick, MA, USA).

Results: At first, 23 independent variables entered to study. After correlation coefficient and regression, 12 variables with $P \leq 0.01$ in Spearman and $P \leq 0.05$ in Pearson were selected. We had the best value of R, MSE and also regression graph in train, validation and tested by Bayesian regularization algorithm with 10 neuron in hidden layer and two delay.

Conclusion: This study showed that artificial neural network (ANN) had acceptable function to extract knowledge from TB raw data; ANN is beneficial to TB incidence prediction.

Keywords: incidence, Iran, neural networks, retrospective studies, tuberculosis.