مقایسه روش‌های رگرسیون خطی و شبکه عصبی مصنوعی در پیش‌بینی میزان مرگ و میر به عنوان تابعی از دمای هوا (مطالعه موردی: تهران)

دکتر منوچهر فرجزاده[1]* محمد دارند[2]

۱- گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشکده ادبیات و علوم انسانی، دانشگاه تربیت مدرس ۲- گروه جغرافیا، دانشکده ادبیات و علوم انسانی، دانشگاه تربیت مدرس
* نویسنده مسؤول: تهران، پل نصر، دانشگاه تربیت مدرس، صندوق پستی ۴۸۳۸-۱۴۱۵۵ تلفن: ۰۹۱۲۱۷۲۳۱۲۴ نمابر: ۰۲۱۸۸۰۰۶۵۴۴
پست الکترونیک: farajzam@modares.ac.ir

دریافت: ۸۸/۳/۱۸ پذیرش: ۸۸/۸/۱۲

چکیده

مقدمه: *تغییرات فصلی و روزانه مرگ و میر ارتباط مستقیمی با دما دارد. در این تحقیق داده‌های روزانه مرگ و میر و پارامتر دما طی دوره ۲۰۰۵ -۲۰۰۲ مورد استفاده قرار گرفته است.*

روش کار: *برای پردازش داده‌ها روش‌های تعیین ضریب همبستگی پیرسون، رگرسیون خطی ساده، رگرسیون چندجمله‌ای و شبکه‌های عصبی مصنوعی به عنوان یک روش غیر خطی (ANN)استفاده شده است.*

یافته‌ها: *نتایج حاصل از کاربرد و تحلیل همبستگی پیرسون نشانگر همبستگی منفی و معنی‌دار بین پارامتر دما با میانگین ماهانه تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی می‌باشد. ارتباط بین این دو با استفاده از شبکه عصبی و الگوریتم‌های ژنتیکی در مقایسه با روش‌های کلاسیک از جمله رگرسیون خطی و رگرسیون چندجمله‌ای نیز نشان می‌دهد که ترکیب شبکه عصبی مصنوعی با الگوریتم ژنتیک نتیجه بهتری را ارایه می‌کند. به این صورت که بعد از آزمون شبکه با لایه‌های پنهان و ضرایب یادگیری مختلف در حالتی که نمونه‌ها به صورت منظم قرار گرفته دقت مدل افزایش پیدا می‌کند.*

بحث: *با توجه به نتایج حاصله می‌توان گفت که شبکه عصبی به خوبی رابطه غیرخطی بین میانگین ماهانه مرگ و میر را در ارتباط با دمای هوا پیش‌بینی می‌کند. ولی در عین حال با ترکیب الگوریتم ژنتیک و شبکه عصبی سرعت تحلیل و دقت فرآیند افزایش می‌یابد به عبارتی دیگر میزان خطا کاهش می‌یابد.*

گل‌واژگان: *مرگ و میر، تهران، شبکه عصبی مصنوعی، مدل رگرسیون خطی و پولی نومیال، درجه حرارت*

مقدمه

بررسی اثر پارامترهای اقلیمی به ویژه دما روی مرگ و میر انسانی توجه بسیاری از محققین را به خود جلب کرده است و محققین مختلف ارتباط معنی‌داری را بین دما و مرگ و میر برخی از شهرهای جهان نشان داده‌اند. در همین زمینه شاموی و همکاران نشان دادند که دما سهم عمده‌ای در افزایش میزان مرگ و میر شهر لوس‌آنجلس آمریکا دارد (۱). همچنین اثر استرس‌زای دما بر مرگ و میر توسط محققان فراوانی به اثبات رسیده است، به این صورت که هرقدر دما از محدوده آسایش انسان بیشتر فاصله گیرد، میزان استرس بیشتر و در نتیجه تعداد فوت‌شدگان نیز افزایش می‌یابد (۶-۲). موضوع مهمی که در بررسی ارتباط بین مرگ و میر با دما مطرح می‌باشد پیش‌بینی آن است که عموماً بر اساس مدل‌های رگرسیونی صورت می‌گیرد که رویکرد جدید استفاده از روش شبکه‌های عصبی برای پیش‌بینی مسایل مختلف است. شبکه‌های عصبی مصنوعی[1] یکی از موفق‌ترین تکنیک‌های داده‌کاوی است که از طریق یک سری مدل‌های پیوندی الهام گرفته شده از رفتار مغز انسان که دارای قابلیت‌های سودمند برای تحقیقات پزشکی است، می‌تواند رابطه غیر خطی بین پدیده‌ها را پیش‌بینی کند

[1] Artificial Neural Network (ANN)

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

(۷). اگرچه روابط غیر خطی بین پارامترها می‌تواند به وسیله مدل‌های متنوع رگرسیونی مدل‌سازی شود ولی طبیعت و ذات این روابط طوری است که چنین مدل‌هایی نمی‌توانند چنین روابطی را به خوبی تجزیه و تحلیل کنند. شبکه‌های عصبی مصنوعی که اغلب شبکه‌های عصبی خوانده می‌شوند به عنوان یک مدل تشخیص روندهای ناپارامتری و غیر خطی می‌تواند روابط غیر خطی پیچیده یا لایه‌های پنهان بین متغیرهای وابسته و مستقل را بیابد و با دقت بهتری نسبت به روش‌های رگرسیونی عمل نماید (۸). بر عکس، زمانی که بین پارامترهای مورد بررسی رابطه غیر خطی آشکاری وجود نداشته باشد به نظر می‌رسد در مواردی شبکه عصبی نمی‌تواند در مقایسه با رگرسیون نتیجه بهتری را ارایه دهد (۹). همچنین شبکه‌های عصبی کاربردهای زیادی در پیش‌بینی سری‌های زمانی دارند (۱۰).

در سال‌های اخیر استفاده از شبکه عصبی مصنوعی جهت پیش‌بینی میزان مرگ و میر و بیماری‌ها بر مبنای پارامترهای اقلیمی در حال افزایش است (۱۱و ۱۲). یکی از رویکردهای امیدوار کننده جهت پیش‌بینی وضعیت مرگ و میر و بیماری‌ها، شبکه‌های عصبی مصنوعی می‌باشد که جهت تشخیص روندهای غیر خطی استفاده می‌شود. به طور مثال خروجی‌های پیش‌بینی شده شبکه‌های عصبی برای بیماری‌های قلبی- عروقی و سرطان بهتر از مدل‌های رگرسیون عمل می‌کنند (۱۳). به علت تغییرات دما طی ماه‌ها و فصول مختلف سال، بین پارامتر دما با کل مرگ و میر و فوت‌شدگان ناشی از بیماری‌های قلبی- عروقی رابطه غیر خطی وجود دارد. یافته‌های متعدد نشان می‌دهند که استفاده از شبکه‌های عصبی برتری خاصی نسبت به روش‌های آماری کلاسیک نظیر رگرسیون ساده و چند متغیره دارد (۱۶-۱۴). امروزه از روش شبکه‌های عصبی در زمینه‌های تحقیقات مختلف پزشکی نظیر اپیدمیولوژی، رادیولوژی و تشخیص تومور (که این بیماری‌ها دارای ارتباط خطی مشخصی با پارامترهای تأثیرگذار ندارند) استفاده می شود (۲۳ -۱۷).

هدف اصلی این مقاله مقایسه و پیش‌بینی مرگ و میر در ارتباط با پارامتر دمای هوا با استفاده از روش‌های رگرسیون خطی و شبکه عصبی مصنوعی در شهر تهران می‌باشد.

روش کار

منطقه مورد مطالعه شهر تهران می‌باشد که از لحاظ موقعیت مکانی در ۳۵ درجه و ۳۵ دقیقه تا ۳۵ درجه و ۴۸ دقیقه عرض شمالی و ۵۱ درجه و ۱۷ دقیقه تا ۵۱ درجه و ۳۳ دقیقه طول شرقی در کوهپایه‌های جنوبی رشته کوه‌های البرز گسترده شده است. داده‌های مورد استفاده در این پژوهش عبارتند از:

داده‌های مرگ و میر: در این تحقیق داده‌های مربوط به میانگین ماهانه تعداد مرگ و میر به تفکیک نوع بیماری طی دوره آماری ۲۰۰۵ -۲۰۰۲ از سازمان بهشت زهرا تهیه شد.

داده‌های اقلیمی: داده‌های اقلیمی روزانه طی دوره ۴ ساله (۲۰۰۲ تا ۲۰۰۵)، از سازمان هواشناسی کشور استخراج شد. سپس از بین پارامترهای اقلیمی میانگین ماهانه حداقل دما در ۵ ایستگاه هواشناسی مستقر در شهر تهران شامل فرودگاه مهرآباد، چیتگر، ژئوفیزیک، دوشان تپه، اقدسیه استخراج شد. در مطالعه حاضر به منظور در اختیار داشتن یک ایده کلی، مقادیر ایستگاه‌های مذکور میانگین‌گیری و در محاسبات مربوطه وارد شد. علت استفاده از داده‌های حداقل دما به خاطر این موضوع بوده که این عامل نسبت به عوامل هواشناسی دیگر در مرگ و میر دارای تأثیر بیشتری می‌باشد و سوابق مطالعاتی هم تأییدکننده این موضوع می‌باشد.

در این مطالعه از روش شبکه عصبی استفاده شده است. اساس شبکه‌های عصبی مصنوعی در واقع شبیه‌سازی تفکر مغز انسان از طریق مدل‌سازی سلول‌های عصبی می‌باشد. هر سلول عصبی به عنوان یک واحد پردازشگر عددی عمل می‌کند. نرون‌های شبکه با هم در ارتباط بوده و به صورت موازی کار می‌کنند. یکی از پرکاربردترین مدل‌های شبکه عصبی، مدل پرسپترون چند لایه است که روی الگوریتم پس‌انتشار[2]، پایه‌گذاری شده است اساس آموزش الگوریتم پس انتشار به این صورت است که وزن‌های اولیه ارتباطی به صورت تصادفی انتخاب می‌گردند. با فرض وجود N نمونه آموزشی که هر نمونه n ورودی و ۱ خروجی داشته باشد، بردار ورودی به صورت $X_j=(X_{1j},...,X_{nj})$ و بردار خروجی به صورت $B_j=(B_{1j},...,B_{lj})$ که $1\le j\le N$، تعریف می‌گردد. فرآیند آموزش با استفاده از دو گام زیر اتفاق می‌افتد. ۱- انتشار به جلو[3]: بردار ورودی X_j به لایه ورودی فرستاده می‌شود و یک بردار خروجی $O_j=(O_{1j},...,O_{lj})$ بر اساس وزن‌های جاری $W=(W_{1l},...,W_{nl})$ ایجاد می‌گردد. مقدار O_j با مقدار خروجی واقعی یعنی B_j مقایسه شده و تابع خطای E تولید می‌شود.

$$E=\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{N}(B_{ij}-O_{ij})^2 \quad (۱)$$

[2] Back propagation (BP)
[3] Forward propagation

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

۲- انتشار خطا به عقب[4]: در این گام، خطای به دست آمده از معادله ۱ به وسیله رابطه زیر روی وزن‌ها توزیع می‌گردد.

$$\Delta W_{nl} = -\frac{\partial E}{\partial W_{nl}}\eta \quad (۲)$$

ضریب $0 < \eta < 1$ پارامتری است که همگرایی نرخ الگوریتم را کنترل می‌نماید. این دو مرحله آنقدر ادامه می‌یابد تا مقدار E به مقدار کوچک از پیش تعیین شده‌ای همگرا گردد. در حقیقت شبکه، با توزیع خطای به دست‌آمده (در هر مرحله) روی وزن‌های قبلی و تکرار این عملیات، آموزش می‌بیند. به تعداد این تکرارها اصطلاحاً اپاک[5] و به روشی که خطاها را به طور معکوس (رو به عقب) روی وزن‌ها توزیع می‌نماید دلتا بار دلتا[6] گفته می‌شود. الگوریتم BP توضیح داده شده در بالا برای مسایل واقعی اغلب بسیار کند است. به منظور غلبه بر این مشکل از الگوریتم ژنتیک[7] برای انتخاب بهترین وزن‌های اولیه BP استفاده می‌گردد (۲۴).

یکی از روش‌هایی که برای بهینه‌سازی چند متغیری استفاده می‌شود، الگوریتم ژنتیک با شبکه‌های عصبی است. شکل مدرن الگوریتم ژنتیک در سال ۱۹۶۰ به وجود آمد. الگوریتم ژنتیک توانسته است محدوده وسیعی از مسایل بهینه‌سازی پیچیده را فقط با استفاده از سه عملیات ژنتیکی ساده (انتخاب، تقاطع و جهش) و بر اساس استفاده از پارامترهای کدبندی شده به جای خود پارامترها، حل نماید. الگوریتم ژنتیک به چندین نکته در فضای تحقیق به طور همزمان توجه می‌نماید و شانس همگرایی به یک بهینه موضعی را کاهش می‌دهد (۲۵و ۲۶).

مزیت این روش این است که علاوه بر زمان اجرای کم با یک بار اجرای برنامه می‌توان به تمام نقاط بهینه رسید و کار طراح در تصمیم‌گیری بسیار آسان می‌شود. روش کار به این صورت است که ابتدا جمعیتی به صورت کاملاً تصادفی ایجاد می‌شود. این جمعیت، شامل تعدادی کروموزوم می‌باشد که هر کروموزوم یک جواب مسأله است و با کدبرداری نمودن آن می‌توان مقدار هر متغیر را بدست آورد. این جمعیت به دو بخش تقسیم می‌گردد، بخش ثابت و بخش متغیر. سپس از این جمعیت، جمعیت اولیه کوچکتری به صورت تصادفی انتخاب می‌گردد. جمعیت انتخاب شده بر اساس تابع برازندگی رتبه‌بندی می‌شود و براساس رتبه، کروموزوم‌های برتر انتخاب می‌گردند. همچنین یکی از کروموزوم‌های برتر بدون تغییر به مرحله بعد (نسل بعد) راه می‌یابد که به این عمل نخبه‌گرایی می‌گویند. سپس جمعیت جدید ایجاد می‌شود و حلقه داخلی فلوچارت تا رسیدن به معیار همگرایی که می‌تواند طی تعداد مشخصی نسل باشد، تکرار می‌گردد.

در این تحقیق برای انجام مطالعه ابتدا از لحاظ زمانی پارامتر دما و مرگ و میر با همدیگر به صورت سری زمانی مرتب گردید. سپس میانگین ماهانه تعداد کل فوت‌شدگان و مرگ و میر ناشی از بیماری‌های قلبی- عروقی مشخص گردید و با داده‌های ماهانه اقلیمی حداقل دما با استفاده از روش‌های آماری ضریب همبستگی پیرسون، معادله خط رگرسیون، رگرسیون چند جمله‌ای به عنوان یک روش نیمه خطی و شبکه‌های عصبی مصنوعی مورد تجزیه و تحلیل قرار گرفت. همچنین جهت پیش‌بینی دقیق مرگ و میر از شبکه عصبی مصنوعی و الگوریتم ژنتیک استفاده شد. در این مقاله ابتدا کل داده‌های مورد استفاده در روش شبکه عصبی به دو قسمت مجزا تقسیم شد که نیمی از داده‌ها برای آموزش شبکه و نیمی دیگر جهت آزمون داده‌ها اختصاص داده شد. تابع تانکسون[8] که تابع معروفی در روش پس انتشار است، به عنوان تابع تحریک استفاده شده است. این تابع بازه هر نرون را میان دو مقدار ۱- و ۱ که همان ضریب همبستگی می‌باشد، فشرده می‌نماید. ضریب همبستگی، بیان کننده مقداری است که میزان رابطه میان خروجی شبکه با مقدار واقعی را مشخص می‌نماید. این ضریب از ۱- تا ۱ تغییر می‌نماید. مقدار ۱ برای ضریب همبستگی به این معنا است که خروجی با مقدار واقعی کاملاً تطابق دارد و ۱- به این معنا است که خروجی و مقدار واقعی کاملاً معکوس هم عمل می‌نمایند. شبکه‌ای با قابلیت مناسب جهت پیش‌بینی مسایل غیرخطی فراهم می‌آورد. برای آموزش شبکه عصبی از نرم‌افزار نرو سولوشن[9] استفاده شده است. تعداد پارامتر دما یک بار به عنوان پارامتر ورودی و پارامتر میانگین ماهانه تعداد کل مرگ و میر به عنوان خروجی در نظر گرفته شدند و بار دیگر پارامتر دما به عنوان ورودی و میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی به عنوان خروجی مورد بررسی قرار گرفت. به منظور کاهش خطای پیش‌بینی، سعی و خطا روی پارامترهای شبکه انجام شد. برای آموزش شبکه از مدل پرسپترون چندلایه[10] استفاده شد که توصیف این مدل به شرح زیر می‌باشد: تعداد لایه‌های مخفی برای آموزش شبکه برابر با ۱ لایه و روش

[4] Error Back propagation
[5] Epoch
[6] Delta Bar Delta
[7] Genetic Algorithm (GA)
[8] TanhAxon
[9] NeuroSolutions
[10] Multilayer perceptron (MLP)

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

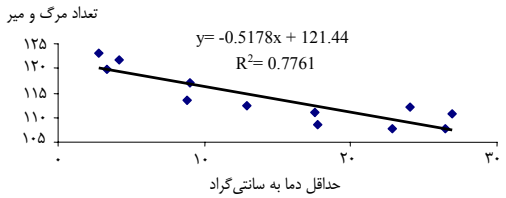
آموزش مورد استفاده دلتا بار دلتا با ضریب یادگیری ۰/۱ می‌باشد که سرعت تطبیق وزن‌های میان نرون‌ها را کنترل می‌نماید یا به عبارتی در واقع طول گام‌ها را در هر بار اصلاح پارامترهای شبکه مشخص می‌کند. از تابع تانکسون، به عنوان تابع تحریک استفاده شد. تعداد نرون‌های لایه مخفی برای آموزش شبکه برابر با ۴ و ضریب مومنتوم که یکی از پارامترهای اصلی آموزش است که نسبت تأثیر وزن‌های قبلی به وزن‌های جدید را مشخص می‌نماید برابر با ۷/. در نظر گرفته شد. تعداد اپاکها برابر با ۱۰۰۰ می باشد. یک مدل شبکه عصبی سه لایه‌ای، شامل لایه ورودی، لایه خروجی و لایه مخفی ایجاد گردید.

نتایج

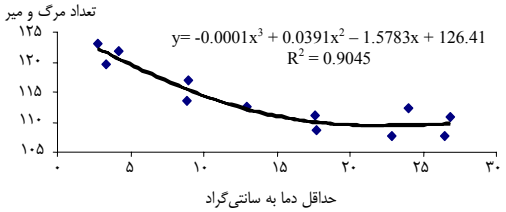
تعداد کل فوت‌شدگان طی دوره آماری ۲۰۰۵ -۲۰۰۲ در شهر تهران برابر با ۱۶۶۰۶۹ نفر می‌باشد. تعداد کل فوت‌شدگان ناشی از بیماری قلبی- عروقی ۳۸/۵٪ (۶۳۹۶۲ نفر) کل تعداد فوت‌شدگان را به خود اختصاص می‌دهد. در تحلیل روش رگرسیونی و نیمه خطی چند جمله‌ای همانطور که در شکل ۱ قسمت الف ملاحظه می‌شود ضریب همبستگی خطی بین تعداد کل مرگ و میر و حداقل دما طی دوره آماری ۲۰۰۵ -۲۰۰۲ برابر با ۰/۸۸- می‌باشد که در سطح ۰/۰۱ =α معنی‌دار بوده و ضریب تعیین آن ۰/۷۷٪ محاسبه شده است. تحلیل همبستگی مذکور نشان می‌دهد که بین میانگین ماهانه کل مرگ و میر با حداقل دما ارتباط معکوس وجود دارد. در روش نیمه خطی چند جمله‌ای مرتبه سوم این اثرات تا ۰/۹۰٪ افزایش یافته است (شکل ۱ قسمت ب).

ضریب همبستگی خطی بین تعداد فوت‌شدگان ناشی از بیماری‌های قلبی- عروقی و حداقل دما طی دوره مذکور برابر با ۰/۹۴- می‌باشد و ضریب تعیین آن برابر با ۰/۸۸ می‌باشد (شکل ۱ قسمت ج). در روش نیمه خطی چند جمله‌ای مرتبه چهارم این اثرات (ضریب تعیین) تا ۰/۹۰٪ افزایش یافته است (شکل ۱ قسمت د). این نتایج نشان می‌دهد که بین تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی با حداقل دما رابطه غیر خطی وجود دارد به صورتی که روش نیمه خطی چند جمله‌ای بهتر و دقیق‌تر از روش رگرسیونی می‌تواند رابطه بین این دو را نشان دهد.

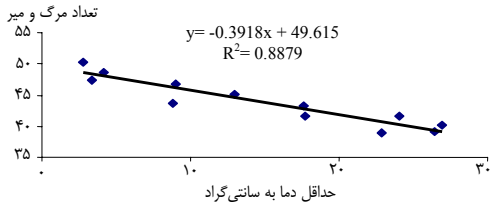
در روش شبکه‌های عصبی مصنوعی نتایج آموزش شبکه با نمونه‌هایی که به صورت منظم قرارگرفته‌اند (منظور از نمونه‌های منظم، افزایش و یا کاهش منظم پارامترهای نمونه است) نشان می‌دهد که در این حالت شبکه پس از آموزش نسبت به حالتی



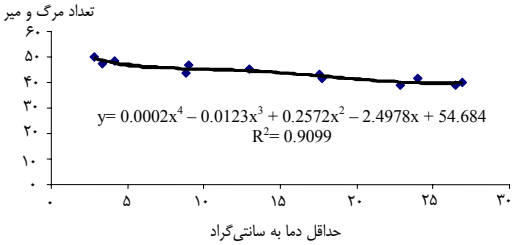
شکل الف- ضریب همبستگی خطی بین میانگین ماهانه تعداد کل مرگ و میر وحداقل دما



شکل ب- همبستگی چند جمله‌ای بین میانگین ماهانه تعداد کل مرگ و میر و حداقل دما



شکل ج- همبستگی خطی بین میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی- عروقی و حداقل دما



شکل د- همبستگی چند جمله‌ای بین میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی- عروقی

شکل ۱- همبستگی خطی و چند جمله‌ای بین میانگین ماهانه تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی با حداقل دما طی دوره آماری ۲۰۰۵ -۲۰۰۲

که در آن نمونه‌ها به صورت تصادفی انتخاب می‌شوند نتیجه بهتری را ارایه می دهد (شکل ۲). همانطور که اشاره شد الگوریتم پس انتشار (BP) برای مسایل واقعی اغلب بسیار کند عمل می‌کند. جهت غلبه بر این مشکل از الگوریتم ژنتیک برای انتخاب بهترین وزن‌های اولیه (BP) استفاده می‌گردد. به عبارت دیگر با استفاده از شبکه عصبی و ترکیب آن با الگوریتم ژنتیک می‌توان سریع‌تر به نتیجه رسید و همچنین دقت را افزایش داد.

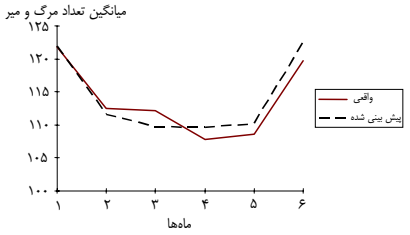
[Downloaded from hakim.tums.ac.ir on 2026-01-26]

شکل ۲ قسمت الف مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر با تعداد پیش‌بینی شده در داده‌های تصادفی و همراه با الگوریتم ژنتیک را نشان می‌دهد. حداقل مجذور متوسط خطاها[11] در این حالت برابر با ۰/۰۱ می‌باشد و ضریب همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر و پیش‌بینی شده توسط شبکه با نمونه‌های تصادفی و همراه با الگوریتم ژنتیک برابر با ۰/۹۴٪ و ضریب تعیین برابر با ۰/۹۰ می‌باشد. شکل ۲ قسمت ب مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر با تعداد پیش‌بینی شده در داده‌های منظم شده و بدون الگوریتم ژنتیک را نشان می‌دهد. حداقل مجذور متوسط خطاها در این حالت برابر با ۰/۰۰۲ می‌باشد و ضریب همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر و پیش‌بینی شده توسط شبکه با نمونه‌های بدون تصادفی و بدون الگوریتم ژنتیک برابر با ۰/۹۶٪ و ضریب تعیین برابر با ۰/۹۳ می‌باشد. شکل ۲ قسمت ج مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر با تعداد پیش‌بینی شده در داده‌های منظم شده و همراه با الگوریتم ژنتیک را نشان می‌دهد حداقل مجذور متوسط خطاها در این حالت برابر با ۰/۰۰۱ می‌باشد و ضریب همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر و پیش‌بینی شده توسط شبکه با نمونه‌های بدون تصادفی و همراه با الگوریتم ژنتیک برابر با ۰/۹۶٪ و ضریب تعیین برابر با ۰/۹۴ می‌باشد.

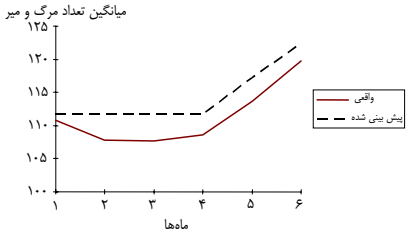
با توجه به موارد فوق می‌توان گفت که آموزش شبکه عصبی مصنوعی در داده‌های منظم شده، همراه با الگوریتم ژنتیک برای پیش‌بینی میانگین ماهانه تعداد کل مرگ و میر در ارتباط با پارامتر حداقل دما نسبت به حالتی که در آن داده‌ها تصادفی و بدون بهره‌گیری از الگوریتم ژنتیک شبکه آموزش داده می‌شود، نتیجه بهتری را ارایه می‌دهد.

شکل ۳ قسمت الف مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده در داده‌های تصادفی و همراه با الگوریتم ژنتیک را نشان می‌دهد. همچنین قسمت ب میزان حداقل مجذور متوسط خطاها را نشان می‌دهد که حداقل مجذور متوسط خطاها برابر با ۰/۰۱ می‌باشد. قسمت ج نیز ضریب همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی و پیش‌بینی شده توسط شبکه با نمونه‌های تصادفی و همراه با الگوریتم ژنتیک را نشان می‌دهد که برابر با ۰/۹۵٪ و ضریب تعیین برابر با ۰/۹۱ می‌باشد.

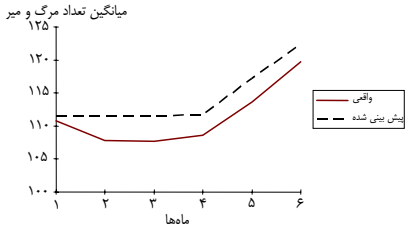
[11] Minimum mean square error (MMSE)



شکل الف- مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر باتعداد پیش‌بینی شده در داده‌های تصادفی و همراه با الگوریتم ژنتیک



شکل ب- مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر باتعداد پیش‌بینی شده در داده‌های منظم شده و بدون الگوریتم ژنتیک



شکل ج- مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر باتعداد پیش‌بینی شده در داده‌های منظم شده و همراه با الگوریتم ژنتیک

شکل ۲- مقایسه میانگین ماهانه تعداد واقعی کل مرگ و میر با تعداد پیش‌بینی شده توسط شبکه در حالت‌های مختلف تصادفی و الگوریتم ژنتیک.

شکل ۳ قسمت د مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده در داده‌های منظم شده و بدون الگوریتم ژنتیک را نشان می‌دهد و قسمت ه میزان حداقل مجذور متوسط خطاها را نشان می‌دهد که برابر با ۰/۰۰۱۹ می‌باشد. قسمت و ضریب همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی و پیش‌بینی شده توسط شبکه با نمونه‌های بدون تصادفی و بدون الگوریتم ژنتیک را نشان می‌دهد که برابر با ۰/۹۷٪ و ضریب تعیین برابر با ۰/۹۵ می‌باشد. شکل ۳ قسمت ز مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده در داده‌های منظم شده و همراه با الگوریتم ژنتیک را نشان می‌دهد و قسمت ک میزان حداقل مجذور متوسط خطاها را نشان می‌دهد که برابر با ۰/۰۰۱ می‌باشد. قسمت ل ضریب همبستگی بین

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی و پیش‌بینی شده توسط شبکه با نمونه‌های بدون تصادفی و همراه با الگوریتم ژنتیک را نشان می‌دهد که برابر با ۰/۹۷٪ و ضریب تعیین برابر با ۰/۹۵ می‌باشد.

در مورد پیش‌بینی میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی- عروقی در ارتباط با پارامتر حداقل دما نیز آموزش شبکه با داده‌های منظم شده همراه با الگوریتم ژنتیک نتیجه بهتری را نسبت به حالتی که در آن داده‌های تصادفی و بدون بهره‌گیری از الگوریتم ژنتیک آموزش داده می‌شود ارایه می‌دهد.

مقایسه روش رگرسیون خطی و پولی نومیال با شبکه عصبی مصنوعی همان طور که در شکل ۱ قسمت الف تا د نشان داده شد بیانگر این است که بین میانگین ماهانه تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی با حداقل دما یک رابطه غیرخطی وجود دارد. ضریب تعیین در ارتباط بین میانگین ماهانه تعداد کل مرگ و میر با حداقل دما با روش رگرسیون خطی برابر با ۰/۷۷ و در روش چند جمله‌ای به عنوان یک روش نیمه خطی برابر با ۰/۹۰ می‌باشد در حالی که در روش شبکه‌های عصبی مصنوعی در داده‌های منظم شده در ترکیب با الگوریتم ژنتیک به۰/۹۴ افزایش یافته است.

همچنین ضریب تعیین در ارتباط بین میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی- عروقی با حداقل دما با روش رگرسیون خطی برابر با ۰/۸۸ و در روش چند جمله‌ای برابر با ۰/۹۰ درحالی که در روش شبکه‌های عصبی مصنوعی در داده‌های منظم شده در ترکیب با الگوریتم ژنتیک به ۰/۹۵ افزایش پیدا کرده است. این نشان می‌دهد که پیش‌بینی با شبکه عصبی مصنوعی می‌تواند نتیجه بهتر و دقیق‌تری را در مقایسه با روش‌های رگرسیونی ارایه دهد (جدول ۱).

جدول ۱- مقایسه ضریب تعیین در روش رگرسیون خطی، چند جمله‌ای و شبکه عصبی مصنوعی

	رگرسیون خطی	پولی نومیال	شبکه عصبی
ضریب تعیین در رابطه بین میانگین ماهانه تعداد کل مرگ و میر با حداقل دما	۰/۷۷	۰/۹۰	۰/۹۴
ضریب تعیین در رابطه بین میانگین ماهانه تعداد مرگ و میر ناشی از بیماری‌های قلبی- عروقی با حداقل دما	۰/۸۸	۰/۹۰	۰/۹۵

بحث

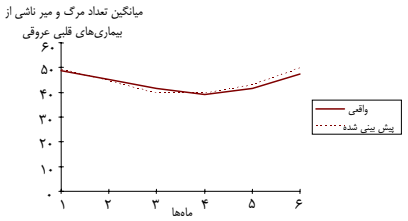
نتایج این مطالعه نشان داد که بین میانگین ماهانه تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی با دمای حداقل ارتباط مستقیم معکوس وجود دارد. این نتیجه در مطالعات متعددی به اثبات رسیده است. ولی استفاده از روابط موجود خطی به علت رفتار پیچیده آن، برای پیش‌بینی مشکل بوده و از این رو نیازمند استفاده از مدل‌هایی است که بتواند بر الگوی روابط موجود بین حداقل دما و مرگ و میر برازش کامل‌تر و بهتری داشته باشد. با توجه به این موضوع این ارتباط که در مدل‌های ساده خطی، خوب می‌باشد، در مدل‌های نیمه خطی چند جمله‌ای بهتر از مدل‌های خطی و در مدل‌های غیرخطی (شبکه عصبی مصنوعی) بسیار قوی می‌باشد و از این رو قادر است دقت پیش‌بینی‌ها را افزایش داده و امکان برنامه‌ریزی‌های لازم را فراهم نماید.

همچنین این مطالعه نشان داد که ارتباط بین میانگین ماهانه تعداد کل مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی با دمای حداقل بیشتر از آن که ساده و خطی باشد، غیرخطی است که توسط شبکه عصبی مصنوعی قابل اندازه‌گیری است (شکل۲). به این صورت که بعد از آزمون شبکه با لایه‌های پنهان و با ضرایب یادگیری مختلف نشان داد که استفاده از شبکه‌های عصبی مصنوعی با یک پرسپترون ۱ لایه پنهان با ضریب یادگیری ۰/۱ و مومنتم ۰/۷ در حالتی که نمونه‌ها به صورت منظم قرار گرفته مدل نسبتاً بهتری را ارایه می‌کند. همچنین بعد از آموزش مجدد شبکه و آزمون شبکه با لایه‌های پنهان و ضرایب مختلف یادگیری در ترکیب با الگوریتم ژنتیک نشان داده شد که ترکیب شبکه با ویژگی‌های مذکور با الگوریتم ژنتیک باعث کاهش خطا و افزایش سرعت محاسبات شده و مدل بهتری را ارایه می‌کند. از این رو استفاده از مدل‌های آزمون شده در این تحقیق قابلیت عملیاتی بیشتری را برای پیش‌بینی آنها داشته و می تواند دقت پیش‌بینی‌ها را افزایش دهد. با توجه به نتایج مثبت و برتری نسبی روش شبکه‌های عصبی در پیش‌بینی مرگ و میر، می‌توان از آن در سایر زمینه‌ها مانند پیش‌بینی مرگ و میرهای ناشی از بیماری‌های گوناگون نیز استفاده نمود.

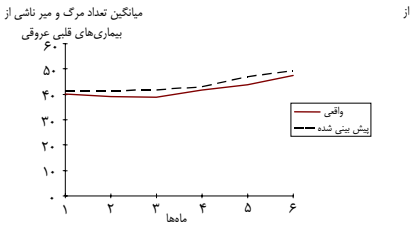
نکته قابل تأمل اینکه در بررسی تحلیل‌های این گونه، وجود آمار دراز مدت احتمالاً نتایج بهتری را به دنبال خواهد داشت و اعتماد نسبت به نتایج حاصله را بیشتر می‌کند. اصولاً توجه به جمع‌آوری دقیق آمار و اطلاعات اهمیت زیادی دارد. در مطالعه حاضر، آماری که توسط بهشت زهرا در اختیار محققین قرار گرفته بود آمار درازمدت‌تری بود که با بررسی‌های انجام شده به دلیل عدم اطمینان به داده‌های قبل از سال ۲۰۰۲ از مطالعات حذف و به تاریخ ذکرشده در مقاله بسنده گردید. در مورد داده‌های هواشناسی نیز این موضوع مصداق دارد. برای مثال

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

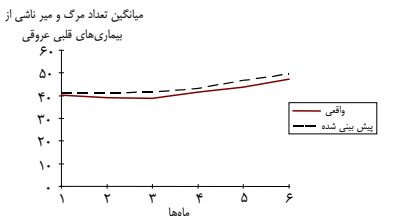
بهتر است در این گونه بررسی‌ها آمار هواشناسی برای هر یک از مناطق ۲۲ گانه شهرداری تهران وجود داشته باشد تا بتوان به برآورد مدل‌ها به تفکیک هر منطقه پرداخت. بدین ترتیب انتظار می‌رود در بین مناطق شهرداری تغییرات فضایی نیز مشاهده گردد که بیان کننده تأثیر شرایط محیطی در هر یک از مناطق است. با توجه به موارد مذکور مشخص است که مطالعه حاضر به عنوان اولین گام‌ها در این زمینه بوده که لازم است مطالعات بعدی تکمیل‌کننده آن باشند تا بتوان به نتایج قطعی‌تر دست یافت.



شکل الف- مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی ازبیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده در داده‌های تصادفی و همراه با الگوریتم ژنتیک

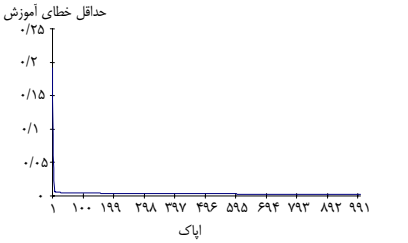


شکل د- مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده در داده‌های منظم شده و بدون الگوریتم ژنتیک

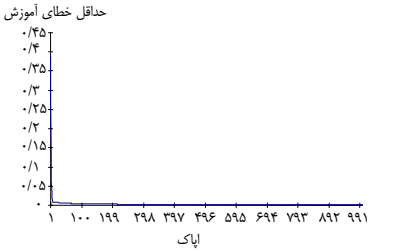


شکل ز- مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده داده‌های منظم شده و همراه با الگوریتم ژنتیک

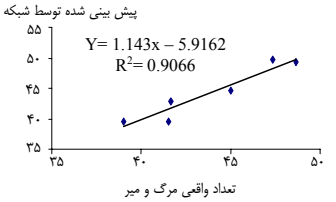
شکل ب- میزان حداقل خطای آموزش شبکه



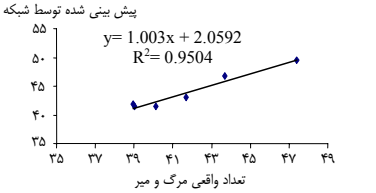
شکل ه- میزان حداقل خطای آموزش شبکه



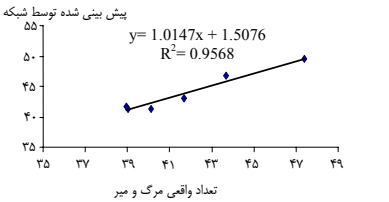
شکل ک- میزان حداقل خطای آموزش شبکه



شکل ج- همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر ناشی ازبیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده توسط شبکه‌های عصبی



شکل و- همبستگی بین میانگین ماهانه تعداد واقعی مرگ و میر ناشی ازبیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده توسط شبکه‌های عصبی



شکل ل- همبستگی میانگین ماهانه بین تعداد واقعی مرگ و میر ناشی ازبیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده توسط شبکه‌های عصبی

شکل ۳- مقایسه میانگین ماهانه تعداد واقعی مرگ و میر ناشی از بیماری‌های قلبی- عروقی با تعداد پیش‌بینی شده توسط شبکه در حالت‌های مختلف تصادفی (تصادفی) و الگوریتم ژنتیک

نتیجه‌گیری

با توجه به نتایج حاصله می‌توان گفت که شبکه عصبی به خوبی رابطه غیرخطی بین میانگین ماهانه مرگ و میر را در ارتباط با دمای هوا پیش‌بینی می‌کند. برای انجام این امر آموزش شبکه با داده‌های مرتب شده برای پیش‌بینی میانگین ماهانه تعداد مرگ و میر و مرگ و میر ناشی از بیماری‌های قلبی- عروقی به علت برازش مناسب داده‌ها دارای دقت بهتری نسبت به داده‌های تصادفی می‌باشد. ولی در عین حال با ترکیب الگوریتم ژنتیک و شبکه عصبی سرعت تحلیل و دقت فرآیند افزایش می‌یابد به عبارتی دیگر میزان خطا کاهش می‌یابد.

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

References

1- Shumway RH, Azari AS, Pawitan Y. Modeling mortality fluctuation in Los Angeles as function of pollution and weather effects. Environ Res 1988; 45: 224-41.

2- Marmor M. Heat wave mortality in New York City, 1949 to 1970. Arch Environ Health 1975; 30: 130-6.

3- Bridger CA, Ellis FP, Taylor HL. Mortality in St Louis during heat waves in 1936, 1953, 1954, 1955 and 1966. Environ Res 1976; 12: 38-48.

4- Ellis FP, Nelson F. Mortality in the elderly in a heat wave in New York City, August 1975. Environ Res 1978; 15: 504-12.

5- Applegate WB, Runyan Jr. JW, Brasfield L, et al. Analysis of 1980 the heat wave in Memphis. J Am Geriatr Soc 1981; 29: 337- 42.

6- Ramlow JM, Kuller LHE. Effects of the summer heat wave of 1988 on daily mortality in Allegheny County, PA. Public Health Rep 1990; 105: 283-9.

7- Haykin S. *Neural networks -a comprehensive foundation*, 2nd ed. New Jersey: Prentice-Hall; 1990.

8- Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet 1995; 346: 1075-9.

9- Namasivayam A, Waldemar A. Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks. Early Hum Dev 2001; 65: 123-37.

10- Lapedes A, Farber R. Non-linear signal processing using neural networks. Technical Report no. LAUR- 2662 Los Alamos National Laboratory, Los Alamos, NM: 1987.

11- Orna I, Nathan I. Interpreting neural-network results: a simulation study. Comput Stat Data Anal 2001; 37: 373-93.

12- Heppard D, Mcphee D. Predicting cytomegalovirus disease after renal transplantation: an artificial neural network approach. Int J Med Inform 1999; 54: 55-76.

13- Jefferson MF, Pendleton N, Lucas SB, et al. comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with non small cell lung carcinoma. Cancer 1997; 79: 1338-42.

14- Yi J, Prybutok VR. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environ Pollut 1996; 92: 349-57.

15- Grander MW, Dorling SR. Neural network modeling and prediction of hourly NOx and NO2 concentrations in urban air in London. Atmos Environ 1999; 33: 709-19.

16- Chaloulakou A, Saisana M, Spyrellis N. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. Science Total Environ 2003; 313: 1-13.

17- Mobley BA, Schechter E, Moore WE, et al. Predictions of coronary artery stenosis by artificial neural network. Artif Intell Med 2001; 18: 187-203.

18- Boone JM. X-ray spectral reconstruction from attenuation data using neural networks. Med Phys 1990; 17: 647-54.

19- Sinha M, Kennedy CS, Ramundo ML. Artificial neural network predicts CT scan abnormalities in pediatric patients with closed head injury. J Trauma 2001; 50: 308.

20- Orr RK. Use of an artificial neural network to quantitate risk of malignancy for abnormal mammograms. Surgery 2001; 129: 459-66.

21- Batuello JT, Gamito EJ, Crawford ED, et al. Artificial neural network model for the assessment of lymph node spread in patients with clinically localized prostate cancer. Urology 2001; 57: 481-5.

22- Henson DB, Spenceley SE, Bull DR. Spatial classification of glaucomatous visual field loss. Br J Ophthalmol 1996; 80: 526-31.

23- Gardner GG, Keating D, Williamson TH, et al. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br J Ophthalmol 1996; 80: 940-4.

24- Sexton RS, Alidaee B, Dorsey RE, et al. Global optimization for artificial neural networks: a tabu search application. Eur J Oper Res 1998; 106: 570–84.

25- Partovi FY, Anandarajan M. Classifying inventory using an artificial neural network approach. Computers and Industrial Engineering 2002; 41: 389–404.

26- Jiang N, Zhao Z, Ren L. Design of structural modular neural networks with genetic algorithm. Advances in Engineering Software 2003; 34: 17–24.

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

Comparing Linear Regression Methods and Artificial Neural Network in Forecasting Human Mortality as a Function of Air Temperature: Case Study of Tehran City

Farajzadeh M*[1] (PhD), Darand M[2] (MS)

[1]*Department of Remote Sensing and GIS, Tarbiat Modares University, Tehran, Iran*
[2]*Departmant of Geography, Tarbiat Modares University, Tehran, Iran*

Received: 8 Jun 2009, Accepted: 3 Nov 2009

Abstract

Introduction: Seasonal and daily human mortality changes have correlation with air temperature. In this research, daily human mortality data and air temperature during 2002- 2005 has been used.

Methods: For data analysis, Pearson adjusted correlation coefficient, polynomial regression as a semi-linear method and artificial neural network as a non-linear method have been used.

Results: The results of Pearson correlation analysis showed significant negative correlation between air temperature and total human mortality and mortality caused by cardiovascular diseases. Their correlation by artificial neural network and genetic algorithm indicated a better result compared to the classic methods (linear and polynomial regression). After network training with different hidden layers and different stepsizes, it was indicated that the use of artificial neural network with one hidden layer of perceptron results in a better model, in the setting of arranged samples.

Conclusion: Therefore, it can be said that neural network can forecast the nonlinear relation between monthly mortality and air temperature, while the combined model of neural network with genetic algorithms can increase analysis speed and accuracy and therefore decrease errors in calculations.

Key words: Mortality, Tehran, Artificial Neural Network, Regression Analysis, Temperature.

Hakim Research Journal 2009; 12(3): 45- 53.

[Downloaded from hakim.tums.ac.ir on 2026-01-26]

*Corresponding Author: Department of Remote Sensing and GIS, Tarbiat Modares University, P.O. Box 14155-4838, Tehran, Iran. Tel: +98- 912- 1723124, Fax: +98- 21- 88006544, Email: farajzam@modares.ac.ir