

ارزیابی مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه در تخمین داده‌های گم شده جریان روزانه (مطالعه موردی: ایستگاه هیدرومتری سنته - استان کردستان)

محمد عیسی‌زاده، پروا محمدی* و یعقوب دین‌پژوه^۱

(تاریخ دریافت: ۱۳۹۴/۱۰/۲۲؛ تاریخ پذیرش: ۱۳۹۶/۰۱/۱۴)

چکیده

تحلیل‌های آماری و پیش‌بینی داده‌های جریان نقش به‌سزایی در مدیریت و مهندسی سامانه‌های آبی ایفا می‌کند. اساسی‌ترین مشکل تحلیل‌های آماری و پیش‌بینی جریان در کشور ایران، عدم وجود دوره آماری طولانی و نقص داده‌های جریان ایستگاه‌های هیدرومتری موجود می‌باشد. با توجه به موارد ذکر شده در این مطالعه، داده‌های جریان روزانه ایستگاه هیدرومتری سنته واقع در استان کردستان با استفاده از داده‌های ایستگاه‌های هیدرومتری و هواشناسی مجاور این ایستگاه تخمین زده شد. این تخمین به‌ازای شانزده ترکیب مختلف، شامل داده‌های جریان روزانه ایستگاه‌های هیدرومتری صفا خانه و پل آتیان و بارش روزانه ایستگاه باران سنجی سنته صورت گرفت. در این تحقیق تخمین جریان روزانه ایستگاه سنته در هر یک از ماه‌های سال، به‌ازای شانزده ترکیب مختلف و مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه مورد ارزیابی قرار گرفت. عملکرد هر یک از مدل‌ها با شاخص‌های NS ، CC ، $RMSE$ و آماره t -student ارزیابی شد. نتایج حاکی از عملکرد خوب هر دو مدل بود ولی مدل شبکه عصبی مصنوعی در تخمین جریان روزانه اکثر ماه‌های سال بهتر از مدل رگرسیونی عمل کرد. میانگین خطای مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه در ماه‌های سال به‌ترتیب برابر با ۶/۳۱ و ۸/۰۷ مترمکعب بر ثانیه برآورد گردید. لازم به‌ذکر است که شبکه عصبی مصنوعی به‌ازای هر یک از شانزده ترکیب به‌کار گرفته شده، عملکرد بهتری نسبت به مدل رگرسیونی داشت.

واژه‌های کلیدی: تخمین داده گم‌شده، شبکه عصبی، رگرسیون خطی چندگانه، ایستگاه سنته

۱. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

*: مسئول مکاتبات: پست الکترونیکی: mohammadi.parva@yahoo.com

مقدمه

نواقص داده‌ها همیشه در تحلیل‌های آماری مشکل‌ساز بوده‌اند. قبل از آنکه متخصص هیدرولوژی به تجزیه و تحلیل داده‌ها اقدام کند لازم است از کمیت و کیفیت داده‌ها و همچنین کامل بودن سری آماری اطمینان حاصل نماید. بدون ارزیابی صحیح داده‌ها، انجام تحلیل‌های پیچیده آماری نتایج قابل اعتمادی را به دست نخواهد داد. یکی از کاربردهای آمار در هیدرولوژی آن است که بتوان برخی خصوصیات آب و هوایی یا هیدرولوژیکی مناطقی را که دارای داده‌های آماری کم یا اصولاً فاقد آمار هستند تخمین زد، زیرا در بسیاری موارد نمی‌توان اجرای یک پروژه را فقط به دلیل این که داده‌های هیدرولوژیکی دراز مدت وجود ندارد به تعویق انداخت. برای رفع خلأهای داده‌های یک ایستگاه اندازه‌گیری معمولاً از روش‌های آماری و با کمک گرفتن از داده‌های ایستگاه‌های مجاور با تشابه هیدرولوژیکی، کلیماتولوژیکی و فیزیوگرافی نسبتاً یکسان استفاده می‌گردد (۶). در اختیار داشتن اطلاعات و آمار صحیح و قابل قبول در بررسی‌های علمی و مطالعات مهندسی منابع آب، طرح‌های توسعه کشاورزی، آبخیزداری و نظایر آن ضروری می‌باشد. این در حالی است که در کشور ما عموماً آمار موجود به دلایل گوناگون که ناشی از عدم دیدبانی یا مشکوک بودن ارقام است، دارای خلأ گسترده می‌باشد. قبل از به کارگیری این آمار در مطالعات لازم است نقایص آنها را بر طرف و داده‌های گم شده را بازسازی کرد. بنابراین دسترسی به داده‌های کافی و دقیق از یک طرف موجب کوتاه تر شدن مدت مطالعات و از طرف دیگر موجب برآورد دقیق‌تر پارامترهای هدف و کاهش هزینه‌های اجرایی و خسارات‌های بعدی ناشی از اجرای طرح‌ها می‌گردد (۷).

دستورانی و همکاران به مقایسه کارایی روش‌های مختلف جهت بازسازی و تطویل داده‌های هیدرولوژی پرداختند (۳). در این تحقیق ترمیم نواقص داده‌های دبی ماهانه با استفاده از مقادیر ایستگاه‌های مجاور و روش‌های نسبت نرمال، همبستگی بین ایستگاه‌ها و شبکه‌های عصبی مصنوعی انجام شد. نتایج

نشان داد مدل شبکه‌های عصبی مصنوعی از کارایی مناسبی جهت پیش‌بینی داده‌های مفقود شده هیدرولوژی برخوردار بود. ریاد و مانیا برای توسعه مدل بارش-رواناب در ایستگاه آقبالو، عملکرد شبکه عصبی مصنوعی را با مدل رگرسیون خطی چندگانه مقایسه کردند (۲۲). در این تحقیق از داده‌های بارش و رواناب با ۷ روز تأخیر ($t-1, t-2, t-3, t-4, t-5, t-6, t-7$)، و همچنین مقدار بارش روز t ام به‌عنوان ورودی مدل‌ها، جهت برآورد مقدار رواناب روز t ام ($Q(t)$) استفاده شد. نتایج قابلیت بیشتر شبکه عصبی مصنوعی نسبت به رگرسیون خطی چندگانه را در مدل‌سازی رواناب نشان داد. ماچادو و همکاران مقایسه‌ای بین مدل شبکه عصبی مصنوعی و مدل مفهومی IPHMEN، در مقیاس زمانی ماهانه جهت پیش‌بینی بارش-رواناب انجام دادند. مدل‌سازی براساس شش ترکیب مختلف ورودی شامل میانگین ماهانه بارش، تبخیر و تعرق، رواناب همزمان و رواناب با تأخیرهای یک و دو روزه انجام گرفت (۱۸). نتایج این مطالعه نشان داد که شبکه عصبی مصنوعی نسبت به مدل مفهومی IPHMEN از دقت بیشتری در پیش‌بینی جریان برخوردار بود. جوشی و پاتال (۱۵) جهت مدل‌سازی بارش-رواناب از سه الگوریتم مختلف شبکه عصبی مصنوعی (Feed Forward Back Propagation, Radial Basis Function و Generalized Regression Neural Network) و رگرسیون خطی چندگانه استفاده کردند. این شبیه‌سازی جریان در دو مرحله انجام گرفت. مرحله اول، از داده‌های بارش با ۴ روز تأخیر به‌عنوان ورودی استفاده شد و سپس در مرحله دوم مقدار جریان با یک روز تأخیر نیز به ورودی‌های بارش اضافه گردید. نتایج نشان داد که الگوریتم FFBP شبکه عصبی مصنوعی دقت بالاتری را نسبت به مدل رگرسیونی داشت. ذونعمت کرمانی و همکاران اثرات جریان ایستگاه‌های بالا دستی را بر عملکرد مدل‌های شبکه عصبی مصنوعی جهت پیش‌بینی رواناب روزانه ایستگاه پایین دست مطالعه کردند (۲۵). سپس با استفاده از شاخص‌های آماری مختلف نتایج

داده‌های اقلیمی در سه شرایط آب و هوایی مختلف ایران پرداختند (۱۶). نتایج نشان داد اگرچه تکنیک‌های هوش مصنوعی پیچیده و زمانبر بود اما نتایج دقیق‌تری نسبت به روش‌های کلاسیک ارائه کرد. همچنین نتایج حاکی از برتری روش رگرسیون چند متغیره از میان روش‌های کلاسیک بود. گکبولاک و همکاران فرایند بارش - رواناب یک حوضه آبخیز را با استفاده از یک روش آماری و روش شبکه‌های عصبی مصنوعی مدل‌سازی کردند (۱۴). مقدار جریان دو روز گذشته (Q_{t-1} , Q_{t-2}) و مقدار بارش روزانه در روز t و $t-1$ به‌عنوان ورودی مدل‌ها جهت برآورد مقدار جریان روزانه به‌کار گرفته شد. نتایج حاکی از عملکرد بهتر شبکه عصبی مصنوعی در شبیه‌سازی جریان بود. نقدی و همکاران به مقایسه روش‌های مختلف تخمین داده‌های گم شده دبی ماهانه حوضه آبخیز کارون بزرگ پرداختند (۹). از روش‌های شبکه‌های عصبی مصنوعی، نسبت نرمال، محور مختصات (گرافیکی)، رگرسیون خطی ساده، رگرسیون خطی چند متغیره و سری‌های زمانی خود همبسته استفاده کردند. با در نظر گرفتن هر ایستگاه به‌عنوان ایستگاه ناقص (پس از حذف عمدی داده‌ها) و دیگر ایستگاه‌ها به‌عنوان ایستگاه‌های شاهد، کار بازسازی آمار ایستگاه ناقص به طریق حذف اعتباری (حذف عمدی) هر یک از داده‌ها انجام شد. نتایج نشان‌دهنده برتری روش شبکه‌های عصبی مصنوعی در مقایسه با دیگر روش‌ها بود. ساداتی‌نژاد و همکاران کارایی روش رگرسیون خطی فازی در مقایسه با روش‌های رگرسیون خطی ساده، رگرسیون چند متغیره و روش محور مختصاتی و نسبت نرمال را برای بازسازی دبی سالانه ایستگاه‌های هیدرومتری حوضه آبریز کارون بزرگ مورد ارزیابی قرار دادند (۴). نتایج نشان داد روش رگرسیون خطی ساده به‌عنوان روش برتر در بازسازی داده‌های دبی سالانه عمل کرد. جلیلیان و همکاران به بررسی کارایی روش‌های مختلف در بازسازی داده‌های هیدرومتری حوضه آبخیز سفیدرود پرداختند (۲). در این پژوهش از چهار روش رگرسیون تک متغیره، رگرسیون دو متغیره، لانگبین و

شبکه عصبی مصنوعی و رگرسیون خطی چندگانه (Multiple Linear Regression) را مورد بررسی قرار دادند. همچنین در این تحقیق از شش ترکیب مختلف جریان روزانه پنج ایستگاه آب‌سنجی موجود در حوضه به‌عنوان ورودی مدل‌ها جهت شبیه‌سازی جریان ایستگاه ماریون استفاده گردید. مقایسه نتایج دو روش نشان داد که روش شبکه عصبی مصنوعی در پیش‌بینی جریان روزانه از دقت بالاتری برخوردار بود. ایچوری و همکاران مدل‌سازی جریان رودخانه میربک را با استفاده از شبکه‌های عصبی مصنوعی انجام دادند و سپس عملکرد آن را با مدل رگرسیون خطی چندگانه مورد مقایسه قرار دادند (۱۰). داده‌های بارش با ۷ روز تأخیر و مقدار بارش روز t به‌عنوان ورودی و مقدار رواناب روز t به‌عنوان خروجی مدل‌سازی‌ها استفاده شد. براساس نتایج مدل‌سازی‌ها مدل شبکه عصبی مصنوعی پیش‌بینی بهتری نسبت به مدل رگرسیون خطی چندگانه داشت. طایفه نسکیلی و همکاران ارزیابی شبیه‌های مختلف شبکه عصبی مصنوعی و روش‌های زمین‌آمار در تکمیل داده‌های مفقود بارش روزانه را انجام دادند (۵). از شش شبکه عصبی مصنوعی به نام‌های MLP، RNN، RBF، TDRNN و CFNN در مقایسه با روش‌های زمین‌آمار کریجینگ و کوکریجینگ مورد بررسی قرار گرفت. جهت ارزیابی از داده‌های بارش ۱۵ ایستگاه باران سنجی موجود در حوضه کرخه استفاده شد. نتایج نشان داد که مدل MLP در برآورد داده‌های مفقود بارش مؤثرتر از بقیه عمل کرد. کوثری و قرده از شبکه عصبی مصنوعی و الگوریتم منطق فازی برای پیش‌بینی جریان حوضه آبریز رودخانه ساویتری استفاده کردند (۱۷). ورودی مدل‌ها شامل ۱۱ ترکیب مختلف شامل مقدار روزانه بارش، متوسط تبخیر روزانه، متوسط درجه حرارت روزانه و مقدار جریان روزانه بود. مقایسه پیش‌بینی‌ها نشان داد که عملکرد مدل شبکه عصبی مصنوعی کاملاً برتر از مدل منطق فازی بود. حسن‌پور کاشانی و دین‌پژوه با استفاده از یازده روش هوش مصنوعی و روش‌های آماری کلاسیک به تعیین مدل مناسب برای برآورد

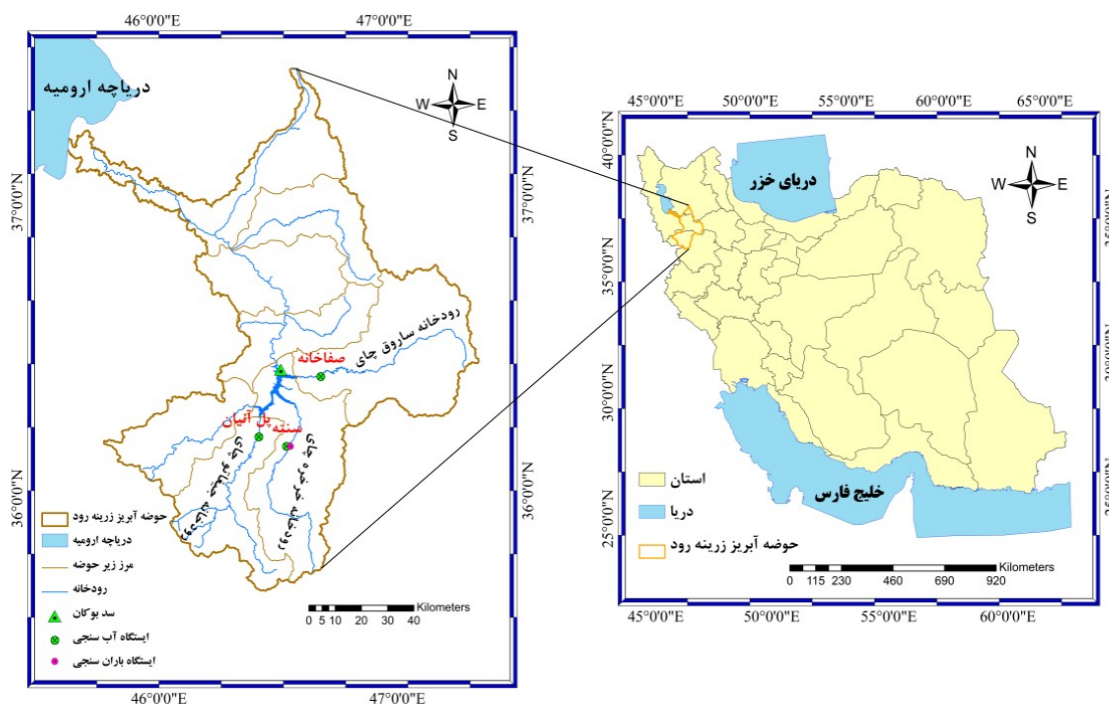
را مطالعه کرد (۲۱). شش ترکیب مختلف داده‌های سه ایستگاه باران سنجی و یک ایستگاه آب‌سنجی با تأخیرهای یک و دو روزه، جهت شبیه‌سازی جریان رودخانه پلاسجان در خروجی حوضه استفاده شد. مقایسه نتایج دو روش نشان داد که روش شبکه عصبی مصنوعی در پیش‌بینی جریان روزانه از دقت بالاتری برخوردار بود. با توجه به پیشینه پژوهش در تحقیق حاضر کارایی مدل شبکه عصبی مصنوعی در کنار مدل رگرسیون خطی چندگانه برای تخمین داده‌های جریان روزانه گم شده حوضه رودخانه خرخره چای مورد ارزیابی قرار گرفت و دقت مدل‌ها در تخمین‌ها نسبت به یکدیگر بررسی گردید. بدین منظور داده‌های جریان روزانه هر یک ماه‌های سال در ایستگاه سنته با استفاده داده‌های بارش و جریان ایستگاه‌های مجاور تخمین زده شد. با توجه به اینکه ترکیب ورودی‌ها و نوع ساختار به‌کار گرفته شده برای هر یک از این مدل‌ها تأثیر چشمگیری در نتایج تخمین و عملکرد آنها دارد؛ بنابراین در این تحقیق تلاش شد کارایی مدل‌های مذکور به‌ازای ۱۶ ترکیب مختلف بارش و جریان ایستگاه‌های مجاور و همچنین انواع ساختارهای مختلف مدل شبکه عصبی مصنوعی مورد ارزیابی قرار گیرد. برای این منظور مدل رگرسیون خطی چندگانه و مدل شبکه عصبی مصنوعی تک لایه با ۱ تا ۲۰ نرون در لایه پنهان جهت تخمین جریان روزانه ایستگاه آب‌سنجی سنته به‌کار گرفته شد.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

در این تحقیق حوضه‌های خرخره چای، ساروق چای و جیغاتو چای که از زیرحوضه‌های اصلی حوضه زرينه‌رود محسوب می‌شوند، به‌عنوان محدوده مورد مطالعه انتخاب گردیدند. این زیر حوضه‌ها در طول جغرافیایی $45^{\circ} 53'$ تا $47^{\circ} 20'$ شرقی و عرض جغرافیایی $35^{\circ} 41'$ تا $36^{\circ} 52'$ شمالی قرار داشت (شکل ۱). این سه حوضه سهم قابل توجهی در حجم آب ورودی به

توماس فایرینگ برای بازسازی داده‌های جریان ماهانه، جریان سالانه، رسوب و کیفیت آب بهره گرفته شد. نتایج به‌دست آمده نشان داد که بین روش‌های مختلف بازسازی دبی متوسط ماهانه در ۱۷ ایستگاه مطالعاتی، از ۱۹۲ مورد بازسازی، روش لانگبین با ۶۶ مورد (۳۳٪) و روش رگرسیون دو متغیره با ۵۲ مورد (۲۷٪) بهترین روش‌های بازسازی جریان متوسط ماهانه بودند، بهترین جواب در روش رگرسیون دو متغیره از بین ۱۷ ایستگاه در ۸ ایستگاه به‌دست آمد و برای برآورد دبی متوسط سالانه مناسب بود. نجفی و همکاران شبیه‌سازی و تحلیل دبی‌های حداکثر لحظه‌ای را با استفاده از شبکه عصبی مصنوعی (مطالعه موردی: ایستگاه‌های هیدرومتری هفت حوض، سولقان، فلاک و مقصود بیک در کلان شهر تهران) انجام دادند (۸). داده‌های مربوط به دبی اوج، دبی ماه بعد و قبل از ماه دبی اوج، دبی حداکثر یک روزه و کد ماه دبی اوج به‌عنوان ورودی شبکه و حداکثر دبی اوج لحظه‌ای به‌عنوان خروجی شبکه تعیین شدند. نتایج تحقیق نشان داد که در همه ایستگاه‌های مورد بررسی، استفاده از شبکه عصبی مصنوعی در شبیه‌سازی دبی‌های حداکثر لحظه‌ای با استفاده از داده‌های دبی حداکثر روزانه و ماهانه از کارایی بالایی برخوردار بود. سلیمانی پیش‌بینی بارش - رواناب را براساس شبکه عصبی مصنوعی (مطالعه موردی: حوضه آبخیز جراحی) بررسی کرد (۲۳). ورودی‌های شبکه عصبی مصنوعی شامل جریان ماهانه، متوسط بارش و دمای هوا برای گام زمانی t و خروجی مدل پیش‌بینی جریان برای گام زمانی $t+1$ بود. نتایج نشان داد پیش‌بینی رواناب رودخانه با مدل شبکه عصبی مصنوعی نسبت به مدل رگرسیون کلاسیک مناسب‌تر و کارآمدتر عمل کرد. الشفیعی و همکاران عملکرد روش‌های شبکه عصبی مصنوعی و رگرسیون را برای پیش‌بینی بارش - رواناب بررسی کردند (۱۳). نتایج نشان داد که شبکه‌های عصبی رفتار رابطه بارش - رواناب با دقت بیشتری نسبت به مدل رگرسیون کلاسیک برآورد کرد. مدرس مدل‌سازی بارش - رواناب با اعتبارسنجی چند معیاره از شبکه عصبی مصنوعی



شکل ۱. موقعیت حوضه آبریز مورد مطالعه و ایستگاه‌های هیدرومتری و باران سنجی

تعیین تعداد ورودی مدل‌ها

در این تحقیق تعداد ورودی‌ها براساس مقدار ضریب همبستگی بین بارش ایستگاه سسته و جریان ایستگاه‌های آب‌سنجی صفا خانه و پل آنیان با جریان ایستگاه سسته تعیین گردید (جدول ۲). نتایج جدول (۲) نشان داد، همبستگی بین بارش ایستگاه سسته با جریان این ایستگاه کم و همبستگی ایستگاه‌های هیدرومتری صفا خانه و پل آنیان زیاد بود. همچنین نتایج حاکی از تغییرات کم ضریب همبستگی در افزایش تعداد تأخیر روزانه بود. با توجه به موارد ذکر شده و سایر مطالعات صورت گرفته در زمینه ترمیم داده‌های گم شده و مدل‌سازی، ترکیب‌های مختلف ورودی‌ها جدول (۳) جهت پیش‌بینی جریان ایستگاه سسته در نظر گرفته شد. در جدول (۳)، $Qs_{anteh}(t)$ نشان‌دهنده جریان ایستگاه آب‌سنجی سسته در زمان t ، $R(t)$ بارش ایستگاه سسته در زمان t ، $Qs(t)$ و $Qs(t-1)$ به ترتیب جریان ایستگاه آب‌سنجی صفا خانه در زمان t و $t-1$ و $Qp(t)$ و $Qp(t-1)$ به ترتیب جریان ایستگاه آب‌سنجی پل آنیان در زمان t و $t-1$ بود.

سد بوکان و در نتیجه تأمین نیازهای آبی پایین دست این سد دارد. ایستگاه‌های سسته، صفا خانه و پل آنیان که به ترتیب در حوضه‌های خرخره چای، ساروق چای و جیغاتو چای واقع شده‌اند از جمله ایستگاه‌های آب‌سنجی هستند که از طول دوره آماری قابل قبول برخوردار می‌باشند. مساحت حوضه بالادست هر یک از ایستگاه‌های آب‌سنجی مذکور به ترتیب بالغ بر ۱۲۳۲، ۲۲۰۹ و ۱۳۲۸ کیلومتر مربع می‌باشند. همچنین متوسط بارش سالانه این حوضه در حدود ۴۰۲ میلی‌متر می‌باشد. داده‌های مورد استفاده در این تحقیق شامل جریان روزانه مشاهداتی در ایستگاه‌های آب‌سنجی سسته، صفا خانه و پل آنیان و بارش روزانه ایستگاه باران سنجی سسته بودند. در این تحقیق از داده‌های روزانه ایستگاه‌های مذکور در دوره آماری (اسفند ۱۳۹۰ - فروردین ۱۳۸۰) جهت مدل‌سازی استفاده شد. مشخصات مربوط به داده‌های جریان و بارش هر یک از ماه‌های سال در جدول (۱) ارائه شده است.

جدول ۱. مشخصات آماری ایستگاه‌های آب‌سنجی و باران سنجی حوضه مورد مطالعه

ماه	باران سنجی سنته			آب‌سنجی صفا خانه			آب‌سنجی پل آنیان		
	ave	max	Std	ave	max	std	ave	max	std
فروردین	۲/۰	۴۰/۰	۵/۵	۲۴/۷	۲۱۱/۲	۲۸/۱	۳۱/۳	۱۵۲/۰	۲۶/۶
اردیبهشت	۱/۸	۶۰/۰	۵/۶	۲۴/۵	۱۶۳/۰	۲۱/۶	۲۳/۴	۱۸۸/۰	۲۱/۲
خرداد	۰/۳	۱۶/۵	۱/۶	۴/۷	۵۶/۶	۷/۵	۵/۶	۳۶/۷	۵/۲
تیر	۰/۰	۶/۰	۰/۳	۰/۹	۴/۴	۱/۰	۱/۰	۷/۴	۱/۰
مرداد	۰/۱	۷/۰	۰/۶	۰/۳	۲/۰	۰/۴	۱/۶	۳۷/۴	۵/۷
شهریور	۰/۱	۱۴/۰	۰/۹	۰/۲	۱۸/۸	۱/۳	۰/۱	۱/۴	۰/۲
مهر	۰/۲	۱۲/۰	۱/۱	۰/۵	۲۰/۳	۱/۴	۰/۲	۲/۹	۰/۴
آبان	۱/۹	۴۰/۰	۵/۲	۲/۱	۲۰/۶	۲/۶	۲/۴	۳۷/۳	۴/۶
آذر	۱/۹	۶۹/۰	۶/۲	۲/۵	۱۳/۶	۱/۷	۳/۲	۴۱/۱	۴/۰
دی	۱/۶	۴۶/۰	۴/۸	۳/۰	۶/۹	۱/۲	۴/۱	۴۷/۵	۴/۷
بهمن	۲/۵	۶۵/۰	۷/۰	۴/۶	۴۲/۵	۳/۸	۷/۹	۷۳/۷	۱۰/۸
اسفند	۲/۰	۴۷/۰	۶/۱	۱۱/۷	۷۳/۸	۱۲/۲	۲۳/۶	۲۴۹/۰	۲۹/۰

توجه: در جدول نمادهای ave, max و std به ترتیب نشان‌دهنده میانگین، حداکثر و انحراف معیار می‌باشد

جدول ۲. ضریب همبستگی بین ورودی‌ها با جریان ایستگاه سنته

تأخیر روزانه	بارش سنته	جریان صفا خانه	جریان پل آنیان
۰	۰/۱۸	۰/۷۱	۰/۶۴
۱	۰/۱۱	۰/۶۹	۰/۶۱
۲	۰/۰۸	۰/۶۸	۰/۵۷
۳	۰/۰۸	۰/۶۸	۰/۵۵
۴	۰/۰۸	۰/۶۸	۰/۵۴
۵	۰/۰۸	۰/۶۹	۰/۵۳
۶	۰/۰۹	۰/۶۹	۰/۵۱
۷	۰/۰۸	۰/۶۸	۰/۵۰
۸	۰/۰۹	۰/۶۸	۰/۴۹
۹	۰/۰۸	۰/۶۸	۰/۴۹
۱۰	۰/۰۹	۰/۶۹	۰/۴۸

شبکه عصبی مصنوعی با آموزش لوئبرگ- مارکوارت

ANN یک سیستم پردازشگر توزیعی اطلاعات است که خصوصیات عملکرد آن شبیه به ساختار عصبی مغز انسان بوده

و گسترش آن براساس قواعد زیر بنا نهاده شده است. ۱- پردازش اطلاعات در عناصر منفردی به نام گره انجام و این گره‌ها در دسته‌هایی به نام لایه مرتب می‌شوند. ANN یک

جدول ۳. ترکیب‌های مختلف بارش و جریان جهت پیش‌بینی جریان ایستگاه آب‌سنجی سنته

شماره ترکیب ورودی	ترکیب ورودی	خروجی
۱	$Qs(t) - R(t)$	$Qsanteh(t)$
۲	$Qp(t) - R(t)$	$Qsanteh(t)$
۳	$Qs(t-1) - Qs(t) - R(t)$	$Qsanteh(t)$
۴	$Qp(t-1), Qp(t), R(t)$	$Qsanteh(t)$
۵	$Qp(t), Qs(t), R(t)$	$Qsanteh(t)$
۶	$Qp(t), Qs(t-1), Qs(t), R(t)$	$Qsanteh(t)$
۷	$Qp(t), Qp(t-1), Qs(t), R(t)$	$Qsanteh(t)$
۸	$Qp(t), Qp(t-1), Qs(t-1), Qs(t), R(t)$	$Qsanteh(t)$
۹	$Qp(t), Qs(t-1), R(t)$	$Qsanteh(t)$
۱۰	$Qp(t-1), Qs(t), R(t)$	$Qsanteh(t)$
۱۱	$Qp(t), Qs(t)$	$Qsanteh(t)$
۱۲	$Qp(t), Qp(t-1)$	$Qsanteh(t)$
۱۳	$Qs(t-1), Qs(t)$	$Qsanteh(t)$
۱۴	$Qs(t-2), Qs(t-1), Qs(t), R(t) Qp(t-1),$	$Qsanteh(t)$
۱۵	$Qs(t-2), Qs(t-1), Qs(t), R(t) Qp(t-1), Qs(t-3),$	$Qsanteh(t)$
۱۶	$Qp(t-1), Qs(t-3), Qs(t-2), Qs(t-1), R(t) Qp(t-2),$	$Qsanteh(t)$

خروجی بستگی به نوع مسئله دارد (تعداد نرون لایه اول برابر تعداد ورودی‌ها و تعداد نرون لایه خروجی برابر تعداد خروجی‌ها می‌باشد) ولی تعداد گره‌های لایه پنهان با سعی خطا به دست می‌آید. در شبکه‌های پیش‌رو، گره‌هایی وجود دارد که در لایه‌های متفاوت قرار دارند. همچنین ورودی‌ها در لایه ورودی شبکه نمایش داده شده و عمل تحریک از ورودی به سمت خروجی شبکه شروع می‌شود (۱۱). لازم به ذکر است که در مطالعه حاضر از تعداد نرون‌های ۱ تا ۲۰ برای لایه پنهان شبکه استفاده گردید تا از این طریق بتوان تأثیر ساختارهای مختلف شبکه بر عملکرد آن در شبیه‌سازی رواناب را مورد ارزیابی قرار داد. از آنجا که متداول‌ترین نوع توابع محرک، تابع تانژانت سیگموئید می‌باشد بنابراین، در این مطالعه از آن برای نگاشت اطلاعات از لایه ورودی به لایه پنهان و از تابع محرک خطی برای نگاشت اطلاعات از لایه پنهان به لایه خروجی استفاده گردید.

سیستم پردازشگر توزیعی اطلاعات است که خصوصیات عملکرد آن شبیه به ساختار عصبی مغز انسان بوده و گسترش آن براساس قواعد زیر بنا نهاده شده است. ۲- پردازش اطلاعات در عناصر منفردی به نام گره انجام و این گره‌ها در دسته‌هایی به نام لایه مرتب می‌شوند. ۳- سیگنال‌ها بین گره‌ها از طریق اتصالات منتقل می‌شود. ۴- هر اتصال وزنی دارد که نشان‌دهنده نیروی اتصال آن است. ۵- هر گره مجموعه وزنی سیگنال‌ها را با به کار بردن یک تبدیل غیرخطی که تابع محرک نامیده می‌شود به سیگنال خروجی تبدیل می‌کند (۱۱ و ۲۴).

حدود ۹۰٪ از مدل‌های شبکه عصبی که در مسائل هیدرولوژی به کار گرفته شده‌اند از نوع شبکه‌های عصبی پیش‌رو هستند (۱۲) که مدل MLP با الگوریتم آموزش لونیگ-مارکواریت از مهم‌ترین آنها به‌شمار می‌رود. معماری معمول این شبکه متشکل از سه بخش لایه ورودی، پنهان و خروجی می‌باشد. تعداد نرون‌های موجود در لایه ورودی و

رگرسیون خطی چندگانه

در رگرسیون چندگانه، مدل شامل یک متغیر وابسته و چندین متغیر مستقل است و معادله رگرسیون به شکل زیر می‌باشد:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

که در آن x_1, x_2, \dots, x_k متغیرهای مستقل، ε میزان اختلاف بین مقدار محاسباتی رگرسیون خطی چندگانه و مقدار واقعی متغیر وابسته، مقدار β_0 عرض از مبدا و β_1 ها ضرایب رگرسیونی یا ضرایب رگرسیونی جزئی را نشان می‌دهند. هر کدام از مقادیر β_1 ها میزان تغییر در مقدار y به ازای یک واحد تغییر در مقدار x متناظر را، وقتی که سایر متغیرهای مستقل ثابت باشند، نشان می‌دهند. مدل رگرسیونی خطی چندگانه برای مشاهده i ام عبارت است از:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad i=1, 2, \dots, n \quad (2)$$

پس از جمع‌آوری مقادیر نمونه‌ای برای هر یک از متغیرهای مستقل و وابسته، یک مدل رگرسیونی چندگانه قابل بسط می‌باشد.

به‌طور کلی، اگر چندین متغیر مستقل در تغییر پذیری متغیر وابسته مؤثر باشند، مدل رگرسیونی چندگانه، گزینه مناسبی برای تحلیل می‌باشد و قادر است متغیر وابسته را با دقت بیشتری پیش‌بینی می‌کند. در رگرسیون خطی چندگانه فرض‌های زیر باید برقرار باشند:

- ۱- خطاها باید دارای توزیع نرمال با میانگین مساوی با صفر و واریانس ثابت باشد: یعنی $\varepsilon_i \sim N(0, \sigma^2)$. ۲- خطاها ناهمبسته باشند. ۳- یک رابطه خطی بین متغیر وابسته و مستقل وجود داشته باشد (فرض خطی بودن). ۴- متغیرهای مستقل با همدیگر همبستگی خطی نداشته باشند (فرض عدم وجود چند هم خطی).

معادله رگرسیونی چندگانه را برای نمونه n تایی با k متغیر مستقل می‌توان به‌صورت زیر نوشت:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \dots + \beta_k x_{2k} + \varepsilon_2 \end{aligned} \quad (3)$$

$y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + \varepsilon_n$
این n معادله را می‌توان به‌صورت ماتریسی زیر نوشت:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (4)$$

معادله فوق را می‌توان به‌صورت ساده زیر نوشت:

$$Y = X\beta + \varepsilon \quad (5)$$

که در آن X ماتریس به ابعاد $n \times (k+1)$ بردارهای Y و ε به ابعاد $n \times 1$ و بردار β به ابعاد $(k+1) \times 1$ است. در این حالت برآورد بردار β (بردار ضرایب رگرسیون خطی چندگانه) به‌روش حداقل مربعات خطا، از رابطه زیر به‌دست می‌آید:

$$\hat{\beta} = (X'X)^{-1} \cdot X'Y \quad (6)$$

X' ترانژاده ماتریس X و $(X'X)^{-1}$ معکوس ماتریس $X'X$ است (۱).

معیارهای ارزیابی

با استفاده از روش‌های مختلفی می‌توان عملکرد مدل‌ها را مورد ارزیابی و مقایسه قرار داد، یکی از این روش‌ها استفاده از معیارهای ارزیابی می‌باشد. از جمله معیارهای ارزیابی پر کاربرد در علوم مهندسی آب، ضرایب، ریشه میانگین مربعات خطا و معیار نش- ساتکلیف که به‌ترتیب در روابط (۷، ۸ و ۹) آورده شدند. دقیق‌ترین مدل با توجه به این معیارها، مدلی خواهد بود که مقدار این سه معیار برای آن به‌ترتیب نزدیک به یک، صفر و یک باشد.

$$CC = \sqrt{\frac{(\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (8)$$

$$NS = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (9)$$

جدول ۴. ارزیابی عملکرد مدل شبکه عصبی مصنوعی در تخمین جریان روزانه ایستگاه سنته در هر یک از ماه‌های سال

ماه	شماره	شماره	آماره‌های ارزیابی عملکرد (آموزش)				آماره‌های ارزیابی عملکرد (صحت‌سنجی)			
			t	NS	RMSE (m ³ /s)	CC	t	NS	RMSE (m ³ /s)	CC
فروردین	۷	۱۴	۸۸/۶۶	۰/۶۱	۸/۹۱	۰/۸۳	۱۹/۴۸	۰/۵۰	۱۸/۸۷	۰/۷۳
اردیبهشت	۱۱	۱۷	۹۶/۰۰	۰/۷۱	۸/۶۲	۰/۸۵	۱۶/۲۸	۰/۳۸	۱۶/۶۴	۰/۶۶
خرداد	۵	۱۳	۸۷/۹۹	۰/۶۷	۱۰/۳۰	۰/۸۲	۱۴/۶۹	۰/۳۲	۴/۲۸	۰/۶۲
تیر	۵	۱۲	۱۰۸/۶۹	۰/۷۶	۸/۶۸	۰/۸۷	۱۸/۲۳	۰/۴۶	۰/۷۳	۰/۷۰
مرداد	۱	۱۳	۷۷/۹۸	۰/۶۲	۱۰/۹۱	۰/۷۹	۲/۴۴	۰/۰۲	۵/۶۷	۰/۱۳
شهریور	۱۲	۱۲	۸۶/۵۳	۰/۶۷	۱۰/۲۵	۰/۸۲	۵/۶۸	-۰/۰۶	۰/۱۸	۰/۲۹
مهر	۱۲	۷	۷۹/۶۵	۰/۶۲	۱۰/۹۶	۰/۸۰	۱/۳۷	-۰/۰۶	۰/۴۳	۰/۰۸
آبان	۸	۱۲	۱۳۶/۳۱	۰/۸۳	۷/۲۳	۰/۹۱	۳۲/۳۰	۰/۷۵	۲/۳۰	۰/۸۷
آذر	۸	۱۸	۱۰۷/۶۱	۰/۷۶	۸/۷۶	۰/۸۷	۳۱/۴۰	۰/۷۳	۲/۰۶	۰/۸۷
دی	۹	۸	۱۰۷/۳۷	۰/۷۶	۸/۷۸	۰/۸۷	۲۸/۳۳	۰/۶۹	۲/۶۰	۰/۸۴
بهمن	۸	۴	۱۱۰/۸۴	۰/۷۷	۸/۴۸	۰/۸۸	۳۲/۲۷	۰/۷۶	۵/۳۲	۰/۸۷
اسفند	۹	۸	۱۱۰/۱۶	۰/۷۷	۷/۲۶	۰/۸۸	۲۵/۸۶	۰/۶۷	۱۶/۵۹	۰/۸۲

در روابط بالا x_i و \bar{x} به ترتیب مقادیر مشاهداتی جریان و میانگین مقادیر مشاهداتی، y_i مقادیر محاسباتی جریان، \bar{y} میانگین مقادیر محاسباتی و N تعداد داده‌ها می‌باشد (۲۰).

آزمون معنی‌داری ضریب همبستگی

پس از تعیین ضریب همبستگی بین داده‌های روزانه مشاهداتی ایستگاه سنته با مقادیر محاسباتی هر یک از مدل‌های منتخب، معنی‌داری ضریب همبستگی هر یک از مدل‌های منتخب توسط آزمون t در سطح معنی‌داری ۰/۰۵ مورد آزمون واقع شد. آماره آزمون t برای بررسی معنی‌داری همبستگی بین متغیرها، به شرح زیر در نظر گرفته شد (۱۹):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (10)$$

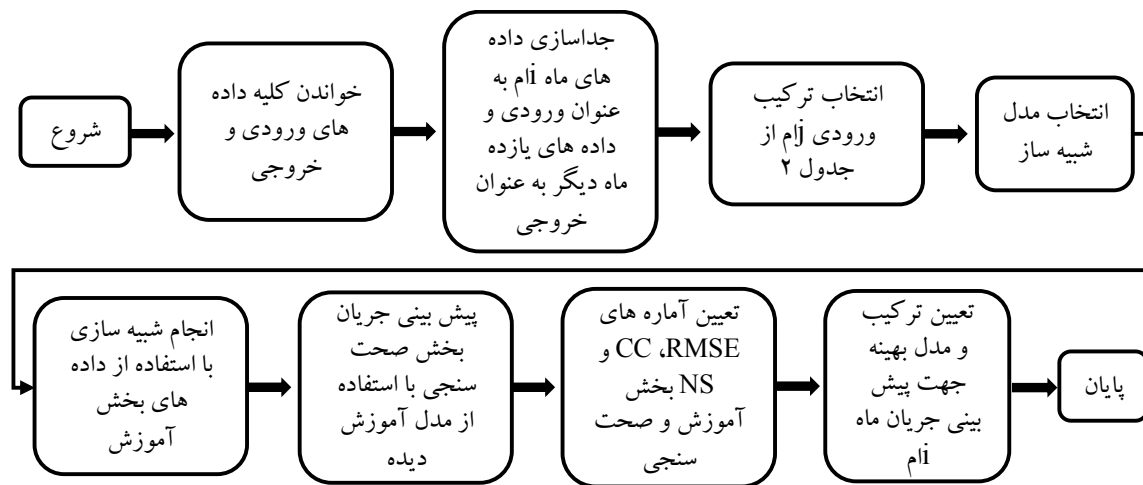
که در آن n تعداد داده‌ها و r ضریب همبستگی به دست آمده بین داده‌های مشاهداتی و محاسباتی مدل‌های منتخب هر ماه می‌باشد که در جدول‌های (۴) و (۵) نشان داده شده است. اگر قدر مطلق آماره t محاسبه شده بیشتر از مقدار t جدول (جدول t -student با درجه آزادی $n-2$ و سطح معنی‌داری $\alpha/2$) گردد، آنگاه ضریب همبستگی مذکور معنی‌دار در نظر گرفته می‌شود.

مراحل پیش‌بینی در هر یک از مدل‌ها

در تحقیق حاضر جهت پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سنته در هر یک از ماه‌های سال، ابتدا کلیه داده‌های بارش ایستگاه سنته و جریان ایستگاه‌های آب‌سنجی صفا خانه و پل آبیان به غیر از ماه مورد نظر (مثلاً ماه فروردین) به عنوان داده‌های ورودی بخش آموزش و داده‌های ایستگاه آب‌سنجی سنته متناظر با ورودی به عنوان خروجی بخش آموزش در نظر گرفته شد. سایر داده‌های جریان و بارش روزانه (مثلاً داده‌های ماه فروردین در سال‌های ۸۰ الی ۹۰) به عنوان بخش صحت‌سنجی در نظر گرفته شد. آنگاه داده‌های بخش آموزش به ازای ۱۶ ترکیب مختلف جدول (۳) با استفاده از مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چند متغیره مورد آموزش قرار گرفتند. در مرحله بعد، پیش‌بینی جریان روزانه بخش صحت‌سنجی با استفاده از مدل‌های آموزش دیده و به ازای ورودی‌های مربوط به بخش صحت‌سنجی انجام گرفت. سپس مقادیر شاخص‌های آماری ذکر شده در بخش قبل، جهت ارزیابی عملکرد مدل‌ها در مراحل آموزش و صحت‌سنجی محاسبه شد. در ادامه با استفاده از این شاخص‌ها بهترین مدل و ترکیب ورودی برای پیش‌بینی جریان روزانه هر یک از ماه‌های

جدول ۵. ارزیابی عملکرد مدل رگرسیون خطی چندگانه در تخمین جریان روزانه ایستگاه سننه در هر یک از ماه‌های سال

ماه	شماره ترکیب منتخب	شماره ترکیب منتخب	آماره‌های ارزیابی عملکرد (آموزش)				آماره‌های ارزیابی عملکرد (صحت‌سنجی)			
			t	NS	RMSE (m ³ /s)	CC	t	NS	RMSE (m ³ /s)	CC
فروردین	۸	۸	۶۳/۹۴	۰/۵۳	۹/۸۶	۰/۷۳	۱۴/۱۳	۰/۲۸	۲۲/۵۱	۰/۶۱
اردیبهشت	۱۳	۱۳	۶۴/۳۲	۰/۵۳	۱۱/۰۴	۰/۷۳	۱۰/۷۲	-۰/۰۱	۲۱/۳۲	۰/۵۰
خرداد	۱	۱	۶۱/۲۰	۰/۵۰	۱۲/۵۷	۰/۷۱	۱۴/۶۴	-۰/۰۸	۵/۳۷	۰/۶۲
تیر	۸	۸	۶۶/۴۴	۰/۵۵	۱۱/۹۹	۰/۷۴	۱۹/۳۷	-۲/۱۷	۱/۷۹	۰/۷۲
مرداد	۹	۹	۶۷/۲۱	۰/۵۴	۱۱/۸۷	۰/۷۴	۲/۷۹	۰/۰	۵/۷۰	۰/۱۵
شهریور	۸	۸	۶۶/۲۱	۰/۵۴	۱۱/۹۸	۰/۷۴	۱۱/۲۰	-۱۶۶/۹۴	۲/۲۱	۰/۵۲
مهر	۸	۸	۶۶/۴۰	۰/۵۴	۱۱/۹۶	۰/۷۴	۰/۰۱	-۳۰/۲۷	۲/۳۵	۰/۰
آبان	۸	۸	۶۶/۸۴	۰/۵۵	۱۱/۹۵	۰/۷۴	۲۷/۶۸	۰/۵۲	۳/۱۶	۰/۸۴
آذر	۱۱	۱۱	۶۶/۰۹	۰/۵۴	۱۲/۰۵	۰/۷۴	۳۰/۰۰	۰/۵۲	۲/۷۳	۰/۸۶
دی	۸	۸	۶۷/۲۱	۰/۵۵	۱۱/۹۴	۰/۷۴	۲۶/۳۸	۰/۵۸	۳/۲۴	۰/۸۲
بهمن	۱	۱	۶۱/۶۹	۰/۵۱	۱۲/۳۷	۰/۷۱	۱۸/۱۲	۰/۱۸	۳/۶۰	۰/۷۱
اسفند	۴	۴	۴۷/۱۹	۰/۳۸	۱۱/۸۶	۰/۶۱	۲۰/۳۳	۰/۳۸	۲۲/۸۸	۰/۷۵



شکل ۲. مراحل انجام شبیه‌سازی با استفاده از مدل‌های مورد استفاده

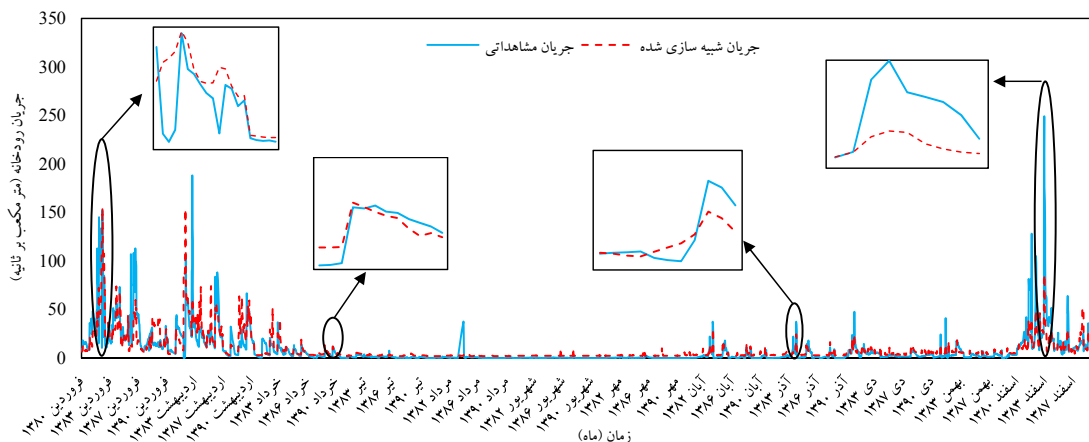
نتایج و بحث

مدل شبکه عصبی مصنوعی

در این مطالعه از شبکه عصبی پرسپترون سه لایه و تعداد نرون ۱ تا ۲۰ جهت پیش‌بینی جریان روزانه ایستگاه سننه در ماه استفاده شد. پیش‌بینی جریان هر یک از ماه‌های سال به‌ازای ۱۶ ترکیب

سال در ایستگاه سننه انتخاب گردید. در شکل (۲) روند انجام شبیه‌سازی جریان به‌ازای مدل‌ها و ترکیب‌های ورودی مختلف نشان داده شده است.

در شکل (۲)، i برابر با ۱ تا ۱۲ (نماد ماه اندازه‌گیری) و j شماره ترکیب ورودی جدول (۳) است.



شکل ۳. مقادیر دبی جریان روزانه مشاهده‌ای و پیش‌بینی شده دوره صحت‌سنجی با مدل‌های منتخب در ایستگاه آب‌سنجی سته

لازم به‌ذکر است که ۷۵ درصد از حجم کل جریان ورودی به ایستگاه سته مربوط به جریان ماه‌های اسفند، فروردین و اردیبهشت بود؛ البته با افزودن درصد جریان ماه‌های بهمن و خرداد این مقدار به ۸۸ درصد رسید. بنابراین مدلی که بتواند جریان این پنج ماه و به خصوص سه ماه اسفند، فروردین و اردیبهشت را با دقت بیشتر و خطای کمتری تخمین بزند، می‌تواند به‌عنوان مدل برتر جهت پیش‌بینی جریان روزانه ایستگاه سته در نظر گرفته شود.

با توجه به اینکه مقدار جدول آماره t-student با درجه آزادی $n-2$ و سطح معنی‌داری ۰/۰۵ تقریباً برابر ۱/۹۷۷ می‌باشد، می‌توان دریافت که مقادیر همبستگی به‌دست آمده همه ماه‌های سال در بخش آموزش جدول (۴)، در سطح معنی‌داری ۰/۰۵ کاملاً معنی‌دار بودند. بررسی معنی‌داری همبستگی دوره صحت‌سنجی هر یک از ماه‌ها به‌ازای آماره t نشان داد که ضریب همبستگی دوره صحت‌سنجی همه ماه‌ها بغیر از ماه مهر معنی‌دار بود. جریان پیش‌بینی شده دوره صحت‌سنجی مدل‌های منتخب هر ماه در شکل (۳) نشان داده شده است.

با توجه به شکل (۳) می‌توان فهمید که مدل شبکه عصبی مصنوعی گاهی اوقات جریان روزانه ایستگاه سته را بیشتر و گاهی کمتر از مقدار واقعی تخمین می‌زند. البته در حالت کلی مقدار جریان تخمین زده شده کمتر از جریان مشاهده‌ای بود.

ارائه شده در جدول (۳) انجام شد. با توجه به معیارهای ارزیابی، ترکیبی که بهترین نتیجه پیش‌بینی را در هر ماه به‌دست می‌داد به‌عنوان بهترین ترکیب جهت پیش‌بینی جریان روزانه آن ماه تعیین گردید. جدول (۴) نتایج بهترین ترکیب ورودی جهت پیش‌بینی جریان هر یک از ماه‌های سال را نشان می‌دهد.

آماره‌های CC ، $RMSE$ و t به‌ترتیب نشان‌دهنده ضریب همبستگی، مجذور میانگین مربعات خطا، شاخص نش-ساتکلیف و آماره t-student محاسباتی است. توجه: برای دیدن فرم ترکیب ورودی با توجه به شماره آن به جدول (۳) مراجعه فرمایید.

با توجه به نتایج درج شده در جدول (۴)، می‌توان نتیجه گرفت که مدل شبکه عصبی مصنوعی می‌تواند مقادیر جریان روزانه ایستگاه آب‌سنجی سته را در ماه‌های فروردین، اردیبهشت، خرداد، تیر، آبان، آذر، دی، بهمن و اسفند به‌خوبی پیش‌بینی کند. با توجه به مقادیر شاخص‌های $RMSE$ و NS دوره صحت‌سنجی این ماه‌ها می‌توان دریافت که استفاده از جریان‌های پیش‌بینی شده در مسائل مدیریتی و ترمیم داده‌های گم شده مطلوب می‌باشد. البته مدل شبکه عصبی مصنوعی در پیش‌بینی جریان روزانه ماه‌های مرداد، شهریور و مهر از دقت قابل قبولی برخوردار نبود. به نظر می‌رسد علت‌های اصلی این امر در کم یا صفر بودن دبی جریان روزانه در این ماه‌ها و همچنین عدم توزیع برابر بارش‌ها در کل سطح حوضه آبریز می‌باشند.

صفا خانه (m^3/s) در روز جاری، $R(t)$ بارش روزانه ایستگاه سنته (mm/day) در روز جاری به میلی متر می باشد.

با توجه به مقادیر شاخص های NS و $RMSE$ دوره صحت سنجی می توان دریافت که استفاده از جریان های پیش بینی شده این شش ماه در مسائل مدیریتی و ترمیم داده های گم شده تا حدودی مطلوب می باشد. البته مدل رگرسیون خطی چندگانه در پیش بینی جریان روزانه ماه های اردیبهشت، خرداد، تیر، مرداد، شهریور و مهر ضعیف عمل کرد، از دلایل این موضوع می توان به کم یا صفر بودن جریان در این ماه ها، عدم توازن نسبی در وقوع بارش ها و همچنین غیرخطی بودن فرایند بارش - رواناب اشاره کرد. با توجه به اینکه مقدار شاخص NS در این ماه ها برابر یا کوچک تر از صفر بود، لذا استفاده از میانگین جریان روزانه بهتر از پیش بینی جریان با مدل رگرسیون خطی چندگانه عمل کرد. مقدار جدول آماره t -student با درجه آزادی $n-2$ و سطح معنی داری 0.05 تقریباً برابر $1/977$ به دست آمد. طبق مقادیر آماره t جدول (۵)، مقادیر همبستگی بخش آموزش به دست آمده در همه ماه ها در سطح معنی داری 0.05 کاملاً معنی دار بودند. همانند نتایج مدل شبکه عصبی در مدل رگرسیونی نیز ضریب همبستگی دوره صحت سنجی ماه مهر غیر معنی دار بود ولی ضریب همبستگی سایر ماه ها در سطح معنی داری 0.05 معنی دار به دست آمد. جریان پیش بینی شده دوره صحت سنجی مدل های منتخب هر ماه با استفاده از رگرسیون خطی چندگانه در شکل (۴) نشان داده شده است.

با توجه به شکل (۴) می توان فهمید که مدل رگرسیون خطی چندگانه نیز مانند مدل شبکه عصبی، گاهی اوقات جریان را بیشتر و گاهی کمتر از مقدار واقعی تخمین می زند. البته در حالت کلی، مقدار جریان تخمین زده شده کمتر از مقدار نظیر جریان مشاهداتی بود و مقدار عددی میانگین از کسر کردن دبی جریان مشاهداتی از مقدار نظیر جریان محاسباتی برابر با (m^3/s) -0.46 به دست آمد. همچنین مدل رگرسیون خطی چندگانه در تخمین مقادیر حداکثر نیز (در برخی مواقع) بسیار

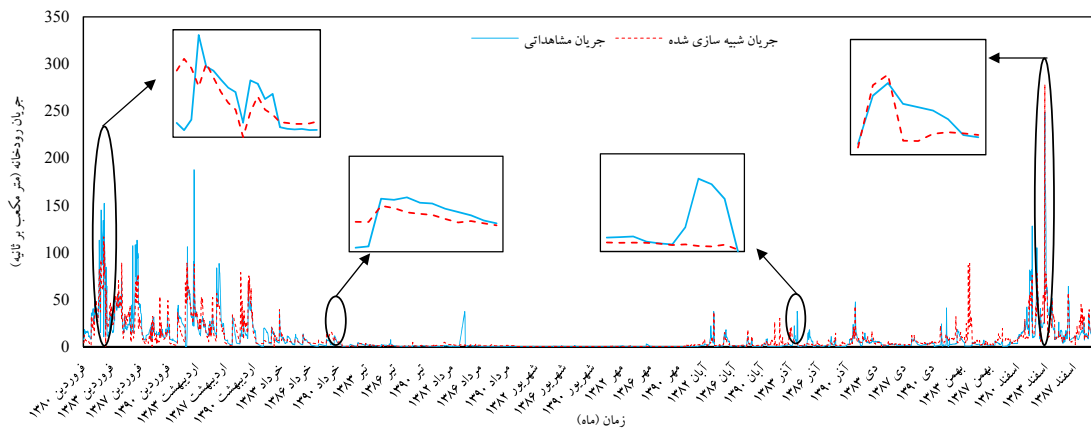
میانگین اختلاف مقادیر به دست آمده از مدل با جریان مشاهداتی برابر (m^3/s) -0.23 حاصل شد. همچنین مدل شبکه عصبی مصنوعی در تخمین مقادیر حداکثر نیز در برخی مواقع بسیار خوب عمل کرد. از عوامل اصلی تغییر در عملکرد مدل می توان به شرایط اولیه رطوبت خاک در هر روز، عدم یکنواختی توزیع بارش در مکان، تغییر در شدت هر بارش در یک منطقه از حوضه اشاره کرد.

مدل رگرسیون خطی چندگانه

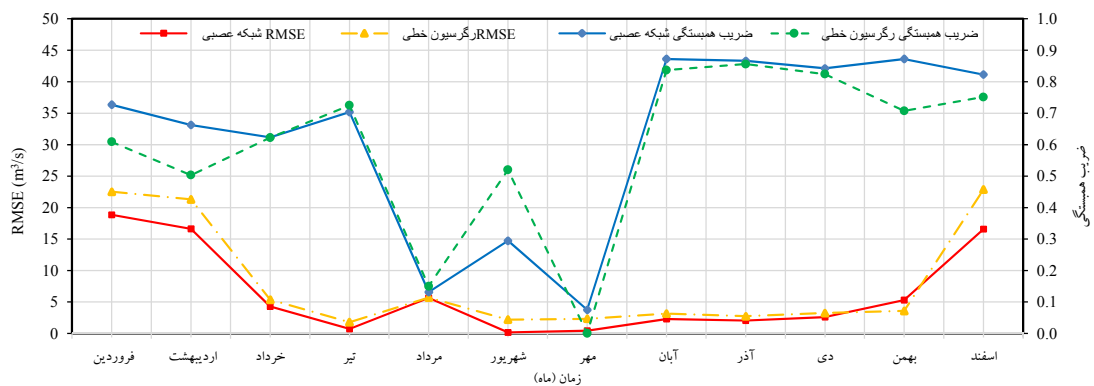
پیش بینی جریان هر یک از ماه های سال به ازای ۱۶ ترکیب ارائه شده در جدول (۳) انجام شده است. با توجه به معیارهای ارزیابی، ترکیبی که بهترین نتیجه پیش بینی را در هر ماه به دست آورده است به عنوان بهترین ترکیب جهت پیش بینی جریان روزانه آن ماه تعیین گردید و در جدول (۵) بهترین ترکیب ورودی مشخص شده است. جدول (۵) نتایج ارزیابی عملکرد مدل رگرسیونی خطی چندگانه را در خصوص بهترین ترکیب ورودی جهت پیش بینی جریان هر یک از ماه های سال نشان می دهد. با توجه به جدول (۵)، می توان نتیجه گرفت که مدل رگرسیون خطی چندگانه قادر است پیش بینی مناسبی از مقادیر دبی جریان روزانه ایستگاه آب سنجی سنته را در ماه های فروردین، آبان، آذر، دی، بهمن و اسفند انجام دهد. به عنوان مثال معادله رگرسیونی منتخب، جهت پیش بینی جریان روزانه ایستگاه سنته در ماه فروردین به ازای پنج ورودی (ترکیب هشتم) به صورت زیر می باشد:

$$Q_{santeh}(t) = 0.25 * Q_p(t) + 0.38 * Q_p(t-1) + 0.07 * Q_s(t-1) + 0.06 * Q_s(t) + 0.22 * R(t) + 1.21 \quad (11)$$

که در آن $Q_{santeh}(t)$ دبی جریان روزانه ایستگاه سنته (m^3/s) در روز جاری، $Q_p(t)$ دبی جریان روزانه ایستگاه پل آبیان (m^3/s) در روز جاری، $Q_p(t-1)$ دبی جریان روزانه ایستگاه پل آبیان (m^3/s) در روز قبل، $Q_s(t-1)$ دبی جریان روزانه ایستگاه صفا خانه (m^3/s) در روز قبل، $Q_s(t)$ دبی جریان روزانه ایستگاه



شکل ۴. نمودار دبی جریان روزانه مشاهده‌ای و پیش‌بینی شده دوره صحت‌سنجی مدل‌های منتخب در ایستگاه آب‌سنجی سنه



شکل ۵. مقایسه عملکرد مدل‌های مورد استفاده با ارائه نمودار مقادیر ضریب همبستگی و RMSE جریان روزانه پیش‌بینی شده در هر ماه در ایستگاه آب‌سنجی سنه

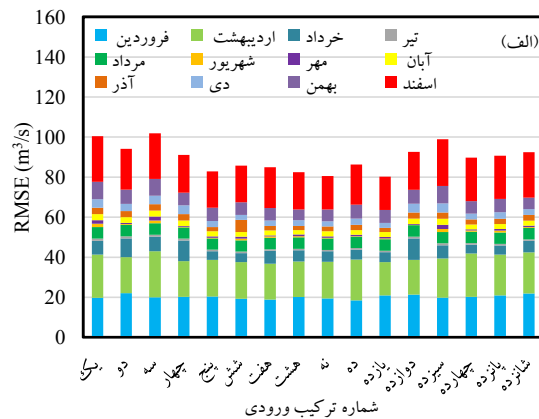
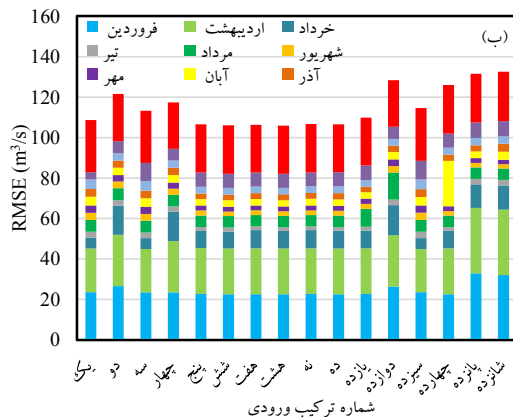
شد. در شکل (۵) مقادیر آماره‌های عملکرد RMSE و CC ترکیب‌های منتخب پیش‌بینی جریان هر یک از ماه‌ها در دوره صحت‌سنجی به‌ازای دو مدل شبکه عصبی مصنوعی و رگرسیون خطی چندگانه نشان داده شده است.

با توجه به شکل (۵) و جدول‌های (۴) و (۵) می‌توان دریافت مقدار RMSE مدل رگرسیون خطی چندگانه در ماه بهمن نسبت به مدل شبکه عصبی مصنوعی $1/72(m^3/s)$ کمتر می‌باشد و همچنین مقادیر ضریب همبستگی جریان در مدل رگرسیون خطی چندگانه در ماه‌های تیر، مرداد و شهریور به ترتیب به اندازه $0/02$ ، $0/02$ و $0/22$ نسبت به مدل شبکه عصبی مصنوعی بیشتر به‌دست آمد. البته لازم به‌ذکر است

خوب عمل نمود. از عوامل اصلی تغییر در عملکرد مدل می‌توان به شرایط اولیه رطوبت خاک در هر روز، عدم یکنواختی توزیع بارش، تغییر در شدت هر بارش و... در هر یک از زیر حوضه‌ها اشاره کرد.

مقایسه عملکرد مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه

با توجه به نتایج به‌دست آمده از مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه در پیش‌بینی جریان هر ماه، در این بخش به بررسی تفاوت عملکرد این مدل‌ها در پیش‌بینی دبی جریان روزانه برای یک‌یک ماه‌های سال پرداخته



شکل ۶. مقادیر RMSE تجمعی دوره صحت‌سنجی به‌ازای شانزده ترکیب ورودی (الف) ANN (ب) رگرسیون خطی چندگانه

ترکیبی که مجموع مقادیر RMSE آن در دوازده ماه سال کمینه به‌دست آمد به‌عنوان بهترین ترکیب معرفی شد. در شکل (۶) مقادیر RMSE هر یک از شانزده ترکیب در هر ماه به‌صورت تجمعی نشان داده شده است.

با توجه به شکل (۶-الف) ترکیب‌های یازده، نه، هشت و پنج به‌ترتیب بهترین ترکیب‌های مورد استفاده جهت پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سننه با شبکه عصبی مصنوعی می‌باشند. همچنین ترکیب‌های هشت، شش، هفت، ده، پنج و نه به‌ترتیب ترکیب‌هایی هستند که نتایج بهتری را در پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سننه با مدل رگرسیون خطی چندگانه به‌دست دادند (شکل ۶-ب). با توجه شکل‌های (۶-الف) و (۶-ب) می‌توان نتیجه گرفت که مدل شبکه عصبی مصنوعی نسبت به مدل رگرسیون خطی چندگانه (به‌ازای کلیه ترکیب‌های ورودی) خطای کمتری (در پیش‌بینی جریان) داشت.

نتیجه‌گیری

در این تحقیق از مدل‌های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی با الگوریتم آموزش لونیبرگ-مارکواریت و تعداد ۱ تا ۲۰ نرون در لایه پنهان جهت پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سننه استفاده شد. پیش‌بینی جریان روزانه در ماه‌های مختلف توسط شانزده ترکیب ورودی (ارائه شده در جدول ۳) با استفاده از مدل‌های مذکور به انجام رسید. شبکه

عملکرد مدل شبکه عصبی مصنوعی بهتر از عملکرد مدل رگرسیون خطی چندگانه بود. میانگین جمع جبری اختلاف RMSE دوره صحت‌سنجی مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندگانه برابر با $1/77 (m^3/s)$ حاصل شد. با توجه جدول‌های (۴) و (۵) مدل شبکه عصبی مصنوعی به‌طور متوسط باعث کاهش ۲۲ درصدی RMSE دوره صحت‌سنجی نسبت به مدل رگرسیون خطی چندگانه شد که این مقدار برابر با ۵۵/۸۲ میلیون متر مکعب آب در مدت یک سال بود. با لحاظ کردن موارد ذکر شده می‌توان به‌صورت قاطعانه مدل شبکه عصبی مصنوعی را بسیار دقیق‌تر و کم‌خطاتر از مدل رگرسیون خطی چندگانه در پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سننه معرفی نمود.

تحلیل اثر ترکیب‌های ورودی در عملکرد مدل‌ها

همان‌طور که اشاره شد در این مطالعه از مدل‌های ANN و MLR جهت پیش‌بینی جریان روزانه هر یک از ماه‌های سال ایستگاه آب‌سنجی سننه، به‌ازای ۱۶ ترکیب مختلف ورودی استفاده شد. بهترین نتایج مربوط به این پیش‌بینی‌ها در جدول‌های (۴) و (۵) درج شد. در این بخش به ارزیابی ۱۶ ترکیب ارزیابی شده پرداخته شد. در واقع مقادیر RMSE دوره صحت‌سنجی هر یک از شانزده ترکیب جدول (۳)، در هر یک از دوازده ماه سال به‌صورت تجمعی بررسی شد. بنابراین،

مدل رگرسیون خطی چندگانه (در پیش‌بینی جریان روزانه) بود، که با نتایج حاصل از این مطالعه نیز همخوانی داشت. استفاده از شانزده ترکیب متمایز در ورودی باعث شد تا مطالعه دقیق‌تر در ارزیابی مدل‌ها و ترکیب‌های مختلف ورودی صورت گیرد. بهترین پیش‌بینی مدل‌ها به‌ازای هر ترکیب برای کلیه ماه‌ها بررسی گردید که نتایج آن حاکی از این بود که شبکه عصبی مصنوعی به‌ازای ترکیب‌های ورودی با شماره یازده، نه، هشت و پنج به‌ترتیب پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سنته را با دقت بالا انجام داد که این ترکیب‌ها شامل داده‌های بارش همزمان، دبی همزمان و روز قبل ایستگاه‌های آب‌سنجی صفا خانه و پل آنیان بودند. مدل رگرسیون خطی چندگانه نیز به‌ازای ترکیب‌های با شماره هشت، شش، هفت، ده، پنج و نه بهترین پیش‌بینی جریان روزانه ایستگاه آب‌سنجی سنته را انجام داد. در مقایسه کلی عملکرد شبکه عصبی مصنوعی جهت پیش‌بینی جریان روزانه کلیه ماه‌ها به‌ازای هر یک از شانزده ترکیب ورودی، از عملکرد مدل رگرسیون خطی چندگانه بهتر عمل کرد. لازم به‌ذکر است که ۷۵ درصد از جریان ایستگاه سنته مربوط به ماه‌های اسفند، فروردین و اردیبهشت بود. در واقع مدلی که بتواند جریان این سه ماه را با دقت بیشتر و خطای کمتری تخمین بزند به‌عنوان مدل برتر جهت پیش‌بینی جریان روزانه ایستگاه سنته برگزیده شد. این مدل برتر طبق نتایج این تحقیق مدل شبکه عصبی مصنوعی بود.

عصبی مصنوعی توانایی زیادی در پیش‌بینی جریان روزانه ماه‌های فروردین، اردیبهشت، خرداد، تیر، آبان، آذر، دی، بهمن و اسفند نشان داد. ولی در پیش‌بینی جریان روزانه ماه‌های مرداد، شهریور و مهر دچار ناتوان بود. از علت‌های اصلی این اختلال می‌توان به وجود تعداد زیاد صفر برای دبی جریان مشاهداتی و همچنین عدم توزیع برابر بارش‌ها در کل سطح حوضه آبریز اشاره نمود. مدل رگرسیون خطی چندگانه نیز در پیش‌بینی جریان روزانه ماه‌های فروردین، آبان، آذر، دی، بهمن و اسفند از دقت خوبی برخوردار بود. این مدل نیز مانند مدل شبکه عصبی در پیش‌بینی جریان روزانه بعضی ماه‌ها عملکرد ضعیفی نشان داد. با مقایسه عملکرد مدل‌ها معلوم شد مدل شبکه عصبی مصنوعی از دقت بیشتر و خطای کمتر (در پیش‌بینی جریان روزانه) برخوردار بود. نتایج حاصل از این پژوهش با پژوهش‌های دستورانی و همکاران (۳) و نقدی و همکاران (۹) در برتری روش شبکه عصبی مصنوعی در مقایسه با روش‌های دیگر در بازسازی داده‌های گم‌شده دبی مطابقت داشت. همچنین نتایج تحقیق ساداتی نژاد و همکاران (۴) در کاربرد روش رگرسیون در برآورد داده‌های گم‌شده دبی با نتایج این پژوهش همخوانی داشت. همچنین نتایج مطالعات ریاد و مانیا (۲۲)، جوشی و پاتال (۱۵)، ذونعمت کرمانی و همکاران (۲۵)، ایچوری و همکاران (۱۰) و نجفی و همکاران (۸) نیز نشان‌دهنده عملکرد بهتر مدل شبکه عصبی مصنوعی نسبت به

منابع مورد استفاده

۱. بیورانی، ح. و ح. جباری خامنه‌ای. ۱۳۹۲. *تحلیل آماری به کمک نرم افزار SPSS و Minitab*. چاپ دوم، انتشارات دانشگاه کردستان، سنندج.
۲. جلیلیان، ح.، ق. رستمی‌زاد و ص. آرخی. ۱۳۹۱. بررسی کارایی روش‌های مختلف در بازسازی داده‌های هیدرومتری (بررسی موردی: حوضه آبخیز سفیدرود). نشریه مرتع و آبخیزداری (مجله منابع طبیعی ایران) ۶۵ (۳): ۳۴۰-۳۲۹.
۳. دستورانی، م. ت.، ع. مقدم‌نیا و ع. طالبی. ۱۳۸۸. بررسی و مقایسه کارایی روش‌های مختلف جهت بازسازی و تطویل داده‌های هیدرولوژی. پنجمین همایش ملی علوم و مهندسی آبخیزداری ایران. دانشگاه علوم کشاورزی و منابع طبیعی گرگان. ۲ اردیبهشت تا ۳ اردیبهشت ۱۳۸۸.
۴. ساداتی‌نژاد، س. ج.، ر. نقدی و م. شایان‌نژاد. ۱۳۸۹. کاربرد روش رگرسیون خطی فازی در برآورد داده‌های ناقص دبی سالانه

- ایستگاه‌های هیدرومتری و مقایسه آن با سایر روش‌های متداول. مجله پژوهش‌های حفاظت آب و خاک ۱۷(۴): ۸۶-۶۷.
۵. طایفه نسکیلی، ن. ب. زهرایی و ب. ثقفیان. ۱۳۹۴. ارزیابی شبیه‌های مختلف شبکه عصبی مصنوعی و روش‌های زمین‌آمار در تکمیل داده‌های مفقود بارش روزانه. مجله مهندسی منابع آب ۸: ۸۷-۶۹.
۶. علیزاده، ا. ۱۳۹۰. اصول هیدرولوژی کاربردی. چاپ سی و سوم، انتشارات دانشگاه امام رضا، مشهد.
۷. لوک‌زاده، صمد. ۱۳۸۳. ارزیابی چند روش مختلف بازسازی خلاء‌های آماری بارندگی در مقاطع زمانی مختلف در منطقه البرز مرکزی. پایان‌نامه کارشناسی ارشد هواشناسی کشاورزی، دانشکده کشاورزی، دانشگاه تهران.
۸. نجفی، ا. ا. صفاری، ع. فنوتای و ا. کرم. ۱۳۹۴. شبیه‌سازی و تحلیل دبی‌های حداکثر لحظه‌ای با استفاده از شبکه عصبی مصنوعی (مطالعه موردی: ایستگاه‌های هیدرومتری هفت حوض، سولقان، قلاک و مقصودبیک در کلان شهر تهران). پژوهش‌های ژئومورفولوژی کمی ۴(۱): ۱۰۳-۹۰.
۹. نقدی، ر. م. شایان نژاد و س. ج. ساداتی‌نژاد. ۱۳۸۹. مقایسه روش‌های مختلف تخمین داده‌های گم شده دبی ماهانه حوضه آبخیز کارون بزرگ. پژوهشنامه مدیریت حوضه آبخیز ۱(۱): ۷۳-۵۹.
10. Aichouri, I., A. Hani, N. Bougherira, L. Djabri, H. Chaffai and S. Lallahem. 2015. River flow model using artificial neural networks. *Energy Procedia* 74: 1007-1014.
11. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial neural networks in hydrology, I: preliminary concepts. *J. Hydrol. Eng.* 5(2): 115-123.
12. Coulibaly, P., F. Anctil and B. Bobée. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* 230: 244-257.
13. El-Shafie, A., M. Mukhlisin, A. A. Najah and R. Taha. 2011. Performance of artificial neural network and regression techniques for rainfall-runoff prediction. *Int. J. Math., Phys. Eng. Sci.* 6(8): 1997-2003.
14. Gökbulak, F., K. Şengönül, Y. Serengil, I. Yurtseven, S. Özhan, H. K. Cigizoglu and B. Uygur. 2015. Comparison of rainfall-runoff relationship modeling using different methods in a forested watershed. *Water Resour Manage.* 29: 4229-4239.
15. Joshi, J. and V. M. Patel. 2011. Rainfall-runoff modeling using artificial neural network (A literature review). National Conference on Recent Trends in Engineering & Technology, 13-14 May, India.
16. Kashani, M. H. and Dinpazhoh. Y. 2012. Evaluation of efficiency of different estimation methods for missing climatological data. *Stoch Environ. Res. Risk Assess.* 26: 59-71
17. Kothari, M. and K. D. Gharde. 2015. Application of ANN and Fuzzy logic algorithms for streamflow modeling of Savitri catchment. *J. Earth Syst. Sci.* 124(5): 933-943.
18. Machado, F., M. Mine, E. Kaviski and H. Fill. 2011. Monthly rainfall-runoff modelling using artificial neural networks. *Hydrol. Sci. J.* 56(3): 349-361.
19. Maidment, D. R. 1993. Handbook of hydrology. New York, McGraw-Hill.
20. Misra, D., T. Oommen, A. Agarwal and S. K. Mishra. 2009. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems Eng.* 103(9): 527-535.
21. Modarres, R. 2009. Multi-criteria validation of artificial neural network rainfall-runoff modeling. *Hydrol. Earth Syst. Sci.* 13: 411-421.
22. Riad, S., J. Mania, L. Bouchaou and Y. Najjar. 2004. Rainfall-runoff model using an artificial neural network approach. *Math Comput Model.* 40: 839- 846.
23. Solaimani, K. 2009. Rainfall-runoff prediction based on artificial neural network (A case study: Jarahi watershed). *Am.-Eurasian J. Agric. Environ. Sci.* 5(6): 856-865.
24. Zealand, C. M., D. H. Burn and S. P. Simonovic. 1999. Short term stream flow forecasting using artificial neural networks. *J. Hydrol.* 214: 32-48.
25. Zounemat-kermani, M., O. Kisi and T. Rajae. 2013. Performance of radial basis and LM-feed forward artificial neural networks for predicting daily watershed runoff. *Applied Soft Computing* 13: 4633-4644.

Evaluation of Artificial Neural Network and Multiple Linear Regression Models to Estimate the Daily Missing data Flow (Runoff) in (Case Study: Santeh Gauging Station- Kordestan Province)

M. Isazadeh, P. Mohammadi* and Y. Dinpazhoh¹

(Received: Aug. 04-2016 ; Accepted: March 07-2017)

Abstract

Statistical analysis and forecast discharge data play an important role in management and development of water systems. The most fundamental issues of statistical analysis and forecast discharge in Iran are lack of data in long term period and lack of stream flow data in gauging stations. Considering the issues mentioned in this study, we tried to estimate the daily data flow (runoff) of Santeh gauging station in Kordestan province using the nearby hydrometric and meteorological stations data. This estimation occurred based on the sixteen different input combinations, including data of daily flow of hydrometric stations Safakhaneh and Polanian and daily runoff in Santeh precipitation gauging station. In this research, the daily flow estimation of the Santeh station in each of the months of the year was evaluated for sixteen different combinations and artificial neural network models and multiple linear regressions. The performance of each model was evaluated with the indicators RMSE, CC, NS and t-student statistic. The results showed good performance of both models but the performance of the artificial neural network model was better than the regression model in estimation of the daily runoff in the most months of the year. Mean error of artificial neural network and multiple linear regression models was respectively estimated as 6.31 and 8.07 m³/s in the months of the year. It should be noted that the artificial neural network, for each sixteen combination used, had better result than the regression model.

Keywords: Artificial Neural Network, Estimating missing data, Multiple linear regression, Santeh station.

1. Dept. of Water Eng., Faculty of Agric., Tabriz Univ., Tabriz. Iran.

*: Corresponding Author, Email: mohammadi.parva@yahoo.com