مقاله کامل علمی- پژوهشی

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و هفتم، شماره پنجم، ۱۳۹۹
۱۴۹-۱۶۶
http://jwsc.gau.ac.ir
DOI: 10.22069/jwsc.2021.17670.3319

رویکرد جدید پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب با استفاده از مدل هیبرید هوشمند (مطالعه موردی: شبکه توزیع آب شهر گرگان)

سید مهران جعفری[۱]، *عبدالرضا ظهیری[۲]، امید بزرگ‌حداد[۳] و محمود محمدرضاپور طبری[۴]

[۱]دانشجوی دکتری سازه‌های آبی، گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان،
[۲]دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، [۳]استاد تمام گروه مهندسی آب، دانشگاه تهران،
[۴]دانشیار گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه مازندران

تاریخ دریافت: ۱۳۹۸/۱۱/۲۰؛ تاریخ پذیرش: ۱۳۹۹/۰۶/۰۵

چکیده

سابقه و هدف: شبکه‌های توزیع آب شهری به‌عنوان یکی از مهم‌ترین تأسیسات و تجهیزات زیربنایی مناطق شهری محسوب می‌شوند. لوله‌ها به‌عنوان یکی از اجزای اصلی و مهم شبکه توزیع آب، همواره در دوره بهره‌برداری تحت‌تأثیر عوامل مختلف دچار شکستگی می‌شوند، بنابراین استفاده از مدل‌های مختلف جهت شناخت و پیش‌بینی نرخ شکست لوله‌ها می‌تواند کاربرد بسیار مهمی برای مدیران و دست‌اندرکاران جهت مدیریت بهینه شبکه توزیع آب شهری در دوران بهره‌برداری داشته باشد. در دهه اخیر مطالعات مختلفی جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب با استفاده از مدل‌های آماری و هوشمند انجام شده است که هر یک دارای نقاط ضعف و قوت می‌باشند. هدف از این پژوهش، ارائه یک رویکرد جدید بر مبنای توسعه مدل پیش‌بینی هیبرید، با توجه به قابلیت‌های مدل‌های هوشمند و آماری، جهت پیش‌بینی دقیق‌تر نرخ شکست لوله‌های شبکه توزیع آب در مقایسه با مدل‌ها آماری و هوشمند مورد استفاده در پژوهش‌های قبل می‌باشد.

مواد و روش‌ها: برای دستیابی به اهداف مطالعه، از آمار دوره زمانی ٤ ساله (۱۳۹٤ تا ۱۳۹۷) مربوط به مشخصات شبکه توزیع آب شرب شهر گرگان شامل قطر، طول، سن، عمق نصب و تعداد شکست جهت پیش‌بینی نرخ شکست لوله‌ها در آینده استفاده شد. برای پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب مورد بررسی، پنج مدل مختلف شامل سه مدل آماری (رگرسیون خطی، رگرسیون خطی تعمیم‌یافته، رگرسیون بردار پشتیبان) و دو مدل هوشمند (شامل شبکه عصبی مصنوعی پیش‌خور و شبکه عصبی مصنوعی پایه شعاعی) مورد بررسی قرار گرفتند. انتخاب پارامترهای بهینه مدل‌های مورداستفاده در این پژوهش، بر اساس شاخص‌های آماری مناسب شامل ضریب همبستگی، خطای حداقل مربعات و نسبت همبستگی- خطای مربعات متوسط مربوط به داده‌های آموزشی و آزمایشی انجام گردید. به‌منظور انتخاب مدل برتر از بین مدل‌های مختلف برای پیش‌بینی نرخ شکست لوله‌های شبکه، مقادیر

* مسئول مکاتبه: zahiri.areza@gmail.com

شاخص‌های R و MSE مدل‌های فوق در مرحله صحت‌سنجی محاسبه و با یکدیگر مقایسه‌گردید. در نهایت، به‌منظور امکان پیش‌بینی دقیق‌تر نرخ شکست لوله‌های شبکه توزیع آب، یک رویکرد جدید بر مبنای مدل پیش‌بینی هیبرید توسعه داده شد که در آن، مقادیر پیش‌بینی‌شده نرخ شکست لوله‌های شبکه توسط هریک از مدل‌های فوق به‌عنوان متغیرهای مستقل ورودی مدل برتر و مقادیر مشاهداتی نرخ شکست به‌عنوان متغیر وابسته خروجی مدل برتر در نظر گرفته شد.

یافته‌ها: مقایسه مقادیر شاخص‌های آماری R و MSE مدل‌های مورد استفاده در این پژوهش در مرحله صحت‌سنجی نشان داد که هیچ‌کدام از مدل‌های مورد استفاده دارای دقت مناسبی برای پیش‌بینی نرخ شکست لوله‌های شبکه شرب شهر گرگان نیستند. مدل شبکه عصبی مصنوعی پیش‌خور با ضریب همبستگی R =۰/٦۹ و مقدار خطای MSE =۰/۰٦۲ دارای بهترین برآورد بود. با استفاده از رویکرد جدید توسعه داده شده بر اساس هیبرید مدل‌های هوشمند و آماری مقدار R برابر ۰/۹٦ و شاخص خطای MSE برابر ۰/۰٤٦ به‌دست آمده است.

نتیجه‌گیری: افزایش چشمگیر شاخص R (به‌میزان ۳۹ درصد) و کاهش قابل‌توجه شاخص MSE (به‌میزان ۲۵ درصد) ناشی از استفاده از رویکرد هیبریدی پیشنهادی برای پیش‌بینی نرخ شکست لوله‌های شبکه در مقایسه با مدل شبکه عصبی مصنوعی پیش‌خور نشان می‌دهد که با استفاده از این رویکرد جدید، می‌توان نرخ شکست لوله‌های شبکه پژوهش حاضر را با دقت بسیار خوبی پیش‌بینی نمود.

واژه‌های کلیدی: شبکه توزیع آب شهری، گرگان، مدل هوشمند، مدل هیبریدی، نرخ شکست لوله

مقدمه

شبکه‌های توزیع آب شهری که جزء اساسی‌ترین سازه‌های زیربنایی مناطق مسکونی به حساب می‌آیند، وظیفه تحویل نیاز آبی شرب، صنعت و بهداشت را برعهده دارند. لوله‌های شبکه توزیع آب در دوران بهره‌برداری به دلایل مختلفی نظیر مشخصات لوله‌ها، شرایط محیطی و بهره‌برداری دچار شکست فیزیکی[1] می‌گردند. بنابراین پیش‌بینی نرخ شکست لوله‌های شبکه در دوران بهره‌برداری می‌تواند چالش‌های پیش‌روی مدیران و بهره‌برداران را برطرف نماید.

در دهه‌های اخیر، پژوهش‌های وسیعی جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب با استفاده از مدل‌های مختلف آماری، هوشمند و هیبرید انجام شده است. استفاده از مدل‌های آماری شامل رگرسیون خطی، مدل آماری نمایی وابسته به زمان، مدل رگرسیون لجستیک جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب توسط پژوهشگران مختلف مانند، شمیر و همکاران (۱۹۷۹)، رجانی و کلینر (۲۰۰۱)، قاسم‌نژاد و همکاران (۲۰۱٤)، فریس و همکاران (۲۰۱٦) و رابلز و همکاران (۲۰۲۰) مورد پژوهش قرار گرفته است (۲۲، ۲۰، ۷، ٦ و ۱۸)، کاربرد مدل‌های شبکه عصبی مصنوعی مصنوعی[2] به‌عنوان یکی از مدل‌های هوشمند و مقایسه آن با مدل‌های دیگر برای مدل‌سازی شکست لوله‌ها و مدیریت نشت شبکه توزیع آب نیز توسط پژوهشگران مختلفی مانند فرناندو (۱۹۹۹)، مانس و همکاران (۲۰۰۲)، جعفر و همکاران (۲۰۱۰) و هو و همکاران (۲۰۱۰)، اثنی‌عشری و همکاران (۲۰۱۳) ستار و

1- Failure

2- Artificial Neural Network (ANN)

همکاران (۲۰۱٦) و کروین و همکاران (۲۰۱۹) مورداستفاده قرار گرفته است (۱٦، ۱۰، ۸، ۲، ۲٦ و ۹). از کاربرد مدل‌های مبتنی بر منطق فازی[1] در زمینه‌های کیفیت آب، نشت آب و شکست لوله‌های شبکه توزیع آب شهری می‌توان به پژوهش‌هایی که توسط صدیق و همکاران (۲۰۰۷)، فارس و همکاران (۲۰۱۰)، اسلام و همکاران (۲۰۱۱)، روهان و همکاران (۲۰۱۷) و پاندی و همکاران (۲۰۲۰) اشاره نمود (۲۲، ۵، ۹، ۱۹ و ۱۷). در زمینه کاربرد هیبرید روش‌های آماری و مدل‌های هوشمند جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب شهری، مدل‌های مختلفی مانند الگوریتم ژنتیک هیبرید[2] و روش رگرسیون تکاملی[3] توسط پژوهشگران مختلفی از جمله کاپلان و همکاران (۲۰۰۳)، براردی و همکاران (۲۰۰۸)، کیانگ یو و همکاران (۲۰۱۱ و ۲۰۱٤)، کاکوداکیز و همکاران در (۲۰۱۷)، فرمانی و همکاران (۲۰۱۷)، طبری و همکاران (۲۰۱۸) و توکلی و همکاران (۲۰۱۹) مورد برررسی قرار گرفته است (۱۲، ۲، ۳۳،۳٤، ۱٤، ٤، ۲۹ و ۳۰).

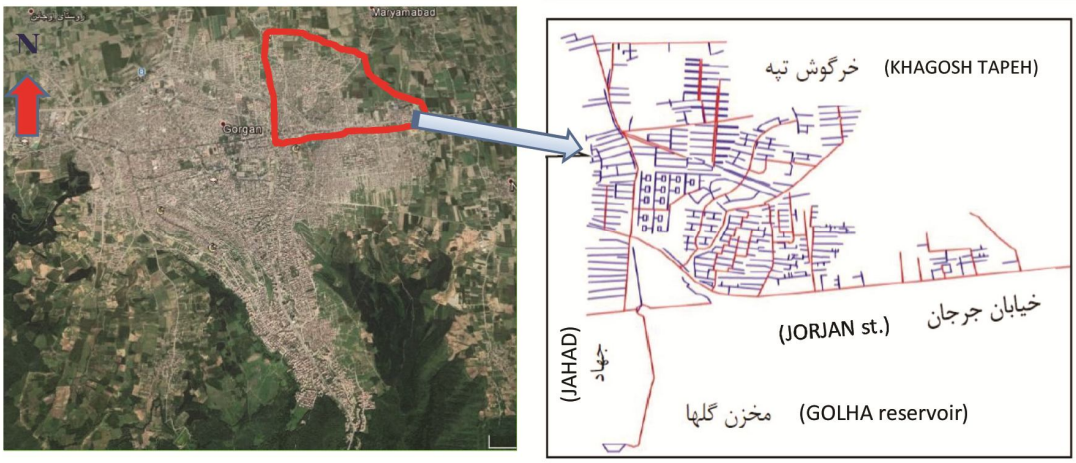
علاوه بر انتخاب مدل مناسب جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب، انتخاب پارامترهای مؤثر بر شکست لوله‌ها نیز بسیار مهم می‌باشد. مشکل اصلی در توسعه مدل‌های پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب، کمبود و یا عدم دسترسی به داده‌های موردنیاز و دقیق بوده به‌طوری‌که ستار و همکاران (۲۰۱٦) در پژوهشی که به‌منظور پیش‌بینی نرخ شکست لوله‌های شبکه انجام دادند، زمان شکست لوله را فقط تابعی از پنج پارامتر قطر، طول، تعداد شکست، حفاظت کاتودیک و پوشش لاینینگ هر لوله درنظر گرفتند (۲٦). هم‌چنین آیدقدو و همکاران (۲۰۱۵) از سه پارامتر طول، قطر و سن لوله به‌عنوان پارامترهای مؤثر بر نرخ شکست لوله‌های شبکه برای پیش‌بینی نرخ شکست لوله‌ها استفاده نمودند (۱). رگرز (۲۰۰۹) به‌منظور پیش‌بینی ریسک شکست لوله و وانگ و همکاران (۲۰۰۹) برای پیش‌بینی نرخ شکست لوله در شبکه توزیع آب پارامترهای قطر، جنس و سن لوله را لحاظ نمودند (۲۱ و ۳۲).

بررسی مطالعات قبلی صورت‌گرفته نشان می‌دهد که استفاده هم‌زمان از قابلیت‌های مدل‌های هوشمند و آماری موجود جهت پیش‌بینی نرخ شکست لوله‌ها مورد توجه قرار نگرفته است؛ بنابراین در این پژوهش، ضمن بررسی توانایی طیف وسیعی از مدل‌های رگرسیون و هوشمند برای پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب شهر گرگان، رویکردی مبتنی بر هیبرید مدل‌های هوشمند در زمینه شکست تجهیزات در شبکه‌های توزیع آب شهری توسعه داده شده است. از شاخص‌های آماری مناسبی برای ارزیابی نتایج مدل‌های توسعه داده شده، استفاده گردید و مدل برتر انتخابی پیشنهاد شد.

مواد و روش‌ها

محدوده مطالعاتی: شهر گرگان طول جغرافیایی ٥٤/۲٦ شرقی و عرض جغرافیایی ۳٦/٥۰ شمالی در استان گلستان و در شمال کشور ایران قرار دارد. آمار حوادث شکست لوله در شبکه توزیع آب شهر گرگان نشان می‌دهد که نرخ شکست لوله‌های شبکه این شهر حدود ۲/۷ (تعداد شکست در هر کیلومتر در سال) بوده که نسبتاً بالا می‌باشد و مشکلات متعددی را در بهره‌برداری از شبکه ایجاد نموده است. در این پژوهش بخش شمال‌شرقی شبکه توزیع آب شهر گرگان به‌دلیل بالا بودن آمار شکست لوله در آن و در نتیجه اهمیت پیش‌بینی نرخ شکست در آینده به‌منظور امکان مدیریت و بهره‌برداری مناسب، به‌عنوان محدوده مطالعاتی انتخاب شده است (شکل ۱).

1- Fuzzy Inference System (FIS)
2- Hybrid Genetic Algorithm (HGA)
3- Evolutional Polynomial Regression



شکل ۱- موقعیت محدوده مطالعاتی در شبکه توزیع آب شهر گرگان.

Figure 1. Location of case study in Gorgan water distribution network.

در پژوهش‌های انجام شده توسط پژوهشگران مختلف به‌منظور و مدلسازی شکست لوله شبکه توزیع آب شهری، پارامترهای مختلفی مانند قطر، طول، جنس، سن، حفاظت کاتودیک و پوشش لاینینگ لوله‌ها درنظر گرفته شده است (۱، ۲۱، ۲۶ و ۳۲)؛ بنابراین در این پژوهش، با توجه به داده‌های موجود و دردسترس، از اطلاعات مربوط به لوله‌های شبکه شهر گرگان در بازه زمانی ٤ ساله (۱۳۹۷-۱۳۹٤) استفاده شده است که شامل قطر (D)، طول (L)، سن (Ag)، عمق کارگذاری لوله‌ها (d) و تعداد شکست هر یک از لوله‌های واقع در محدوده شبکه می‌باشند. براساس آمار موجود شرکت آب و فاضلاب استان گلستان، مشخصات شبکه توزیع آب در محدوده مطالعاتی در جدول ۱ ارائه شده است.

جدول ۱- مشخصات شبکه توزیع آب محدوده مطالعاتی.

Table 1. Water distribution network properties of the case study.

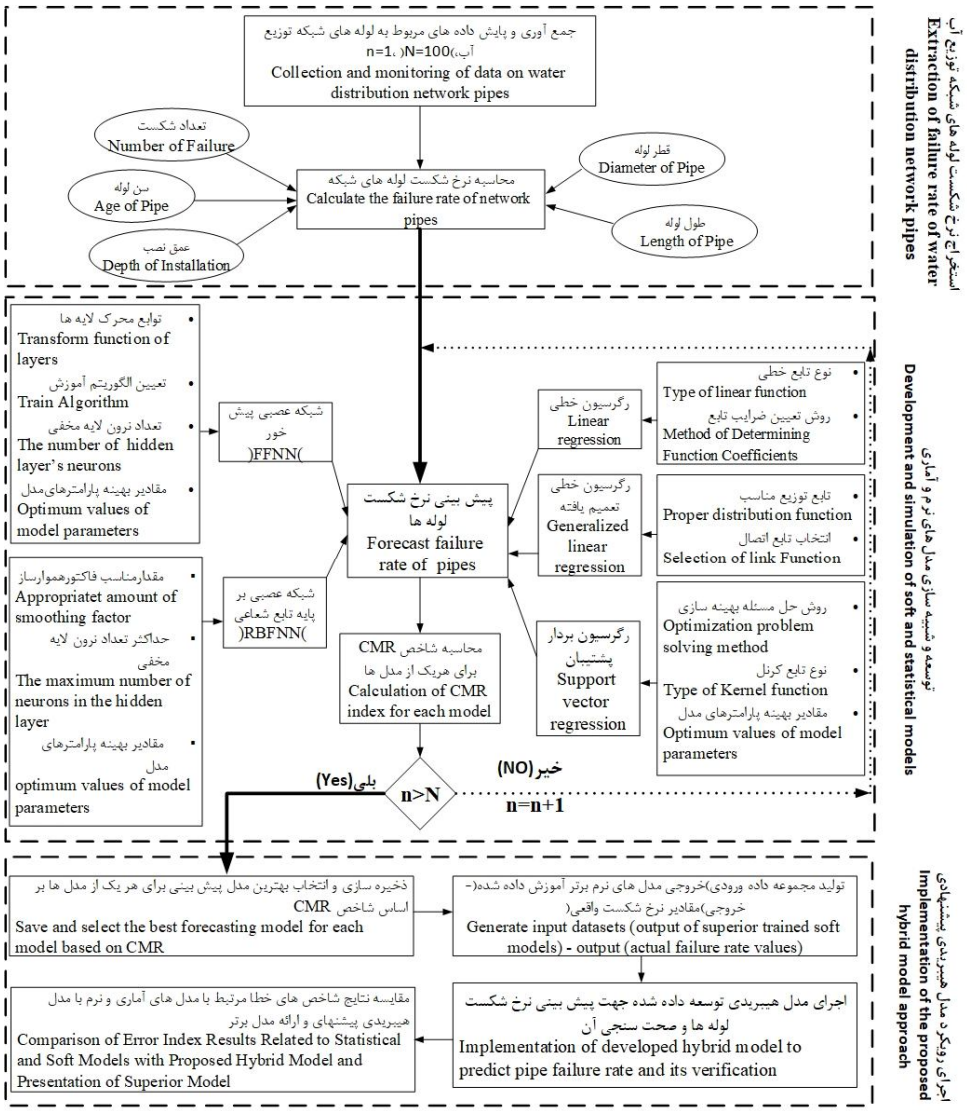
مشخصه Property	مقدار Value
بازه زمانی مورد بررسی (Investigated time duration)	1394-1397
محدوده قطر لوله-میلی‌متر (Rang of pipe diameters-mm)	63-500
طول کل لوله- متر (Total length of pipeline-m)	80072
تغییرات طول لوله (Pipe length changes)	9-134
تعداد کل لوله‌ها (Number of pipes)	1547
تعداد کل شکست‌ها (Number of failures)	169

نرخ شکست[1] هر لوله بر حسب تعداد شکست در واحد طول در هر سال از رابطه ۱ محاسبه گردید:

$$FR = \frac{\mathrm{NF}}{L \times T} \qquad (۱)$$

که در آن، NF تعداد کل شکست لوله طی دوره آماری، L طول لوله و T برابر طول مدت دوره آماری (سال) می‌باشد.

ساختار مدل پیشنهادی توسعه داده شده: در این پژوهش به‌منظور توسعه مدل پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب محدوده مطالعاتی، از هیبرید روش‌های رگرسیون و محاسبات هوشمند استفاده گردید. ساختار مدل پیشنهادی این پژوهش به‌صورت شکل ۲ می‌باشد.



شکل ۲- ساختار رویکرد پیشنهادی جهت پیش‌بینی نرخ شکست لوله‌ها.

Figure 2. Structure of Proposed Approach to Predict Pipes Failure Rate.

1- Failure Rate (FR)

• استخراج نرخ شکست مشاهداتی مربوط به لوله‌های شبکه توزیع آب براساس مشخصات مربوط به لوله‌های شبکه شامل طول، قطر، سن، عمق کارگذاری و تعداد شکست‌های.

• توسعه و شبیه‌سازی بهترین مدل‌های هوشمند و رگرسیون (آماری) با استفاده از داده‌های مرحله قبل و سپس محاسبه مقادیر نرخ شکست توسط بهترین هریک از مدل‌ها

• توسعه و اجرای مدل هیبرید پیشنهادی و انتخاب مدل برتر با استفاده از مقادیر نرخ شکست مرحله قبل و شاخص‌های خطای مناسب.

شاخص‌های ارزیابی مدل: به‌منظور بررسی دقت و مناسب‌بودن مدل‌های برازش داده شده بر اساس داده‌های مشاهداتی، شاخص‌های آماری مختلفی وجود دارند. با توجه به اهمیت دو شاخص ضریب همبستگی[1] (R) و خطای میانگین مربعات[2] (MSE) از بین شاخص‌های موجود، در این پژوهش از این دو شاخص برای بررسی نکویی برازش و دقت مدل جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب استفاده شد. رابطه‌های ۲ و ۳ به‌ترتیب برای محاسبه شاخص‌های R و MSE استفاده می‌شوند (۱):

$$R = \frac{\sum_{i=1}^{n}(x_{mi}-\bar{x}_m)(x_{ci}-\bar{x}_c)}{\sqrt{\sum_{i=1}^{n}(x_{ci}-\bar{x}_c)^2 \sum_{i=1}^{n}(x_{mi}-\bar{x}_m)^2}} \quad (۲)$$

$$MSE = \frac{\sum_{i=1}^{n}(x_{mi}-\bar{x}_m)}{n} \quad (۳)$$

که در آن، x_{mi} مقادیر مشاهداتی، x_{ci} مقادیر پیش‌بینی شده، $\bar{x}_m$ متوسط مقادیر مشاهداتی، $\bar{x}_c$ متوسط مقادیر پیش‌بینی‌شده و n تعداد مشاهدات می‌باشند.

در این پژوهش، از شاخص نسبت خطا[3] که براساس نسبت R به MSE تعریف می‌شود (رابطه ٤)، به‌منظور بررسی عملکرد و دقت مدل پیش‌بینی استفاده شد. جهت جلوگیری از بیش‌برازش[4] و انتخاب مدل مناسب و هم‌چنین انتخاب مقادیر بهینه پارامترهای برخی مدل‌ها در فرآیند مدلسازی، ابتدا کل داده‌های مشاهداتی به سه دسته داده‌های آموزشی، آزمایشی و صحت‌سنجی با نسبت‌های ۷۰، ۱۵ و ۱۵ درصد از کل داده‌ها تقسیم گردید. سپس مدل‌های موردنظر ۱۰۰ بار مورد اجرا قرار گرفته و مجموع شاخص ER داده‌های آموزشی و آزمایشی هریک از اجراها که به‌صورت شاخص نسبت همبستگی-میانگین مربعات خطا[5] (رابطه ۵) تعریف می‌گردد، محاسبه می‌شوند. مقادیر پارامترهای اجرایی از مدل پیش‌بینی که دارای بیش‌ترین مقدار CMR می‌باشد، به‌عنوان مقادیر بهینه و مدل مربوطه بهترین مدل پیش‌بینی برای هر مدل انتخاب می‌گردد.

$$ER_i = \frac{R_i}{MSE_i} \quad (٤)$$

$$CMR_i = ER_i^{Train} + ER_i^{Test} \quad (٥)$$

که در آن‌ها، R_i، MSE_i، ER_i و CMR_i مقادیر شاخص‌های مربوط به هر تکرار برای مدل‌های مورد استفاده در این پژوهش می‌باشند.

علاوه بر دو شاخص فوق، شاخص‌های راندمان نش-ساتکلیف[6] (NS) و میانگین قدرمطلق خطاها[7] (MAE) برای ارزیابی دقت مدل‌ها استفاده می‌گردد. مقدار شاخص NS بین ۱- و ۱ تغییر می‌کند،

1- Coefficient of correlation (R)
2- Mean Square Error (MSE)
3- Error Ratio (ER)
4- Over fitting
5- Correlation Mean square error Ratio (CMR)
6- Nash–Sutcliffe efficiency (NS)
7- Mean Absolute Error (MAE)

به‌طوری‌که چنان‌چه مقدار این شاخص برابر یک باشد نشان‌دهنده تطابق بسیارخوب بین نتایج مدل با مقادیر مشاهداتی است. با توجه به این‌که شاخص MAE نشان‌دهنده اختلاف بین مقادیر پیش‌بینی و مشاهداتی می‌باشد بنابراین هرچه به صفر نزدیک‌تر باشد، مطلوب‌تر است. مقادیر این شاخص‌ها با استفاده از رابطه‌های ٦ و ٧ محاسبه می‌گردند:

$$NS = 1 - \frac{\sum_{i=1}^{n}(X_{ci}-X_{mi})^2}{\sum_{i=1}^{n}(X_{mi}-\bar{X}_m)^2} \qquad (٦)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_{ci} - X_{mi}| \qquad (٧)$$

مدل رگرسیون خطی: رگرسیون خطی یک رویکرد خطی بین متغیر پاسخ با یک یا چند متغیر توصیفی است. در صورت به‌کارگیری چندین متغیر توصیفی یا مستقل در مدل رگرسیونی، روش رگرسیونی را چندگانه می‌گویند (٢). باتوجه به وجود مقادیر حدی در برخی از داده‌های مشاهداتی، به‌منظور کاهش اثرات داده‌های حدی، استفاده از روش حداقل مربعات هوشمند به‌منظور تعیین ضرایب متغیرهای مستقل، جواب بهتری نسبت به سایر روش‌ها ارائه می‌دهد. محاسبه مقادیر مربعات خطا در این روش نیز به دو روش حداقل قدرمطلق باقی‌مانده و مربعات وزنی دوگانه انجام می‌شود. در این روش انتخاب تابع وزنی مناسب از بین توابع وزنی متفاوتی مانند Andrews، Bisqure، Cauchy و غیره برای محاسبه وزن هریک از متغیرها مهم می‌باشد. در این پژوهش با توجه به وجود داده‌های حدی، از روش رگرسیون چندگانه که در آن محاسبه حداقل خطا با روش RLS و مربعات وزنی دوگانه انجام می‌شود، استفاده شده است.

رگرسیون خطی تعمیم‌یافته: در رگرسیون خطی تعمیم‌یافته، می‌توان توابع توزیع مختلفی علاوه بر تابع توزیع نرمال را برای متغیر پاسخ درنظر گرفت (١٣). مدل رگرسیون خطی تعمیم‌یافته از سه جزء شامل، تابع توزیع حاکم بر متغیر وابسته یا پاسخ با میانگین μ، بردار ضرایب b که یک ترکیب خطی Xb از متغیرهای مستقل را برقرار می‌نماید و تابع اتصال که به‌صورت زیر تعریف می‌شود:

$$f(\mu) = Xb \qquad (٨)$$

در این پژوهش تابع توزیع حاکم بر متغیر پاسخ یا وابسته تابع توزیع نوع پواسون می‌باشد و بنابراین تابع اتصال مناسب نیز تابع Log می‌باشد.

مدل رگرسیون بردار پشتیبان: ماشین بردار پشتیبان[1] مجموعه‌ای از نقاط در فضای k بعدی (که k معرف تعداد ویژگی‌های هر داده است) داده‌ها است که مرز دسته را مشخص می‌کند و مرزبندی و دسته‌بندی داده‌ها براساس آن‌ها انجام می‌شود و با جابجایی یکی از آن‌ها ممکن است خروجی دسته‌بندی تغییر نماید. SVR نسخه‌ای از مدل SVM است که به‌جای طبقه‌بندی داده‌ها، عمل رگرسیون را انجام می‌دهد (١٨). در مدل‌سازی یک پدیده با استفاده از روش SVR برای یک مجموعه مشاهداتی دارای N مشاهده که x_k پارامترهای ورودی و y_k پارامتر خروجی می‌باشند، هدف یافتن بهترین تابع خطی به شکل زیر می‌باشد:

$$f(x) = y_k = WX_k^T + b \qquad (٩)$$

برای یافتن بهترین تابع $f(x)$ از بهینه‌سازی با تابع هدف زیر استفاده می‌شود:

$$\text{Minimize}\left\{\frac{1}{2}\,WW^T + \sum_{k=1}^{K}(\xi_k + \xi_k^*)\right\} \qquad (١٠)$$

1- Support Vector Machine (SVM)

که در آن، W بردارهای وزن بهینه، b بردارهای اریب بهینه، ξ_k^*, ξ_k متغیرهای کمکی، C یک عدد مثبت است که کنترل‌کننده میزان جریمه اختصاص داده شده به مشاهدات خارج از حاشیه هوشمند است که از بیش‌برازش مدل جلوگیری می‌کند.

شبکه عصبی مصنوعی پیش‌خور: شبکه عصبی مصنوعی پیش‌خور مرسوم‌ترین ساختار شبکه‌های پس‌انتشار می‌باشد که اغلب دارای یک لایه ورودی، یک یا چند لایه مخفی از نرون‌های زیگموئید و یک لایه خروجی است که دارای تابع خطی می‌باشد. تعیین مشخصات لایه‌های مخفی در کارایی عملکرد شبکه بسیار مهم می‌باشد (۲). این مشخصات شامل ابعاد شبکه یعنی تعداد لایه مخفی و نرون‌های هر لایه، نوع توابع محرک و روش‌های آموزش شبکه است. با مقایسه مقادیر شبیه‌سازی شبکه‌های مختلف با مقادیر واقعی می‌توان مناسب‌ترین ابعاد شبکه را به‌دست آورد. در شبکه عصبی پیش‌خور از تابع فعال‌ساز لجستیک استفاده می‌شود.

شبکه‌های عصبی مصنوعی تابع پایه شعاعی: این شبکه‌ها نوعی از شبکه‌های عصبی مصنوعی نظارت شده هستند که دارای ساختار پیش‌خور بوده و مشتمل بر یک لایه ورودی، یک لایه مخفی و یک لایه خروجی می‌باشند. در این شبکه تعداد نرون‌های لایه مخفی بر اساس پارامترهای ورودی تعیین می‌شوند. رابطه تابع محرک RBF پایه شعاعی به شکل رابطه ۱۱ می‌باشد:

$$a = radbase(n) = e^{-n^2} \qquad (11)$$

که در آن، n بردار فاصله بین بردار وزن‌های w و بردار ورودی P می‌باشد که در مقدار اریب (b) ضرب شده است و a مقدار خروجی تابع RBF است که مقدار بین صفر و یک را به خود می‌گیرد.

لایه خروجی در شبکه RBF دارای تابع محرک خطی Pureline می‌باشد. آن‌چه که برای ساخت یک شبکه عصبی مصنوعی پایه شعاعی لازم است، مشخص کردن مقدار فاکتور هموارساز و تعیین مقدار شاخص خطای موردنظر است.

توسعه مدل پیش‌بینی هیبرید: به‌دلیل پیچیدگی برخی پدیده‌ها و هم‌چنین تأثیر عوامل مختلف بر آن پدیده، عملکرد مدل‌های هوشمند و آماری در پیش‌بینی یک پدیده همواره دارای محدودیت‌هایی می‌باشند که این امر موجب عدم مطابقت نتایج مدل‌سازی با استفاده از مدل‌های هوشمند و آماری با مقادیر مشاهداتی و در نتیجه کم‌بودن دقت مدل‌های پیش‌بینی می‌گردد. در این پژوهش به‌منظور افزایش دقت و عملکرد مدل پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب، مدل هیبرید با استفاده از مزایای سه مدل رگرسیون و دو مدل هوشمند جهت پیش‌بینی نرخ شکست لوله‌ها توسعه داده شد. ساختار این مدل به‌صورت زیر می‌باشد:

- جمع‌آوری اطلاعات مربوط به شبکه توزیع آب موردنظر شامل قطر، سن، طول، عمق نصب و تعداد شکست‌های لوله و دسته‌بندی آن‌ها به داده‌های آموزشی، آزمایشی و صحت‌سنجی.
- آموزش و انتخاب بهترین مدل: مدل‌های آماری و هوشمند مختلف مورداستفاده در این پژوهش، با استفاده از داده‌های آموزشی و آزمایشی، به تعداد ۱۰۰ بار اجرا می‌گردد و سپس بهترین مدل مربوط به هریک از مدل‌ها، براساس مقادیر شاخص CMR انتخاب می‌گردد.
- استخراج ورودی برای مدل هیبرید: با استفاده از بهترین مدل‌های انتخابی از مرحله قبل، نرخ شکست لوله‌های شبکه توزیع آب پیش‌بینی می‌گردد. نرخ شکست پیش‌بینی شده لوله‌ها، به‌عنوان ورودی مدل هیبرید در نظر گرفته می‌شود.

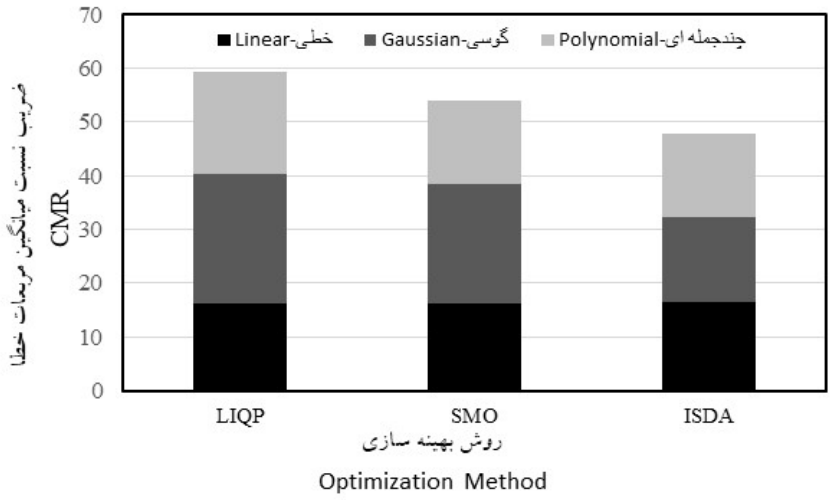
• تعیین مشخصات پارامترهای مدل هیبرید: مدل هیبرید، با استفاده از داده‌های مرحله قبل، به تعداد ۱۰۰ بار اجرا و مقادیر پارامترهای مربوط به بهترین مقادیر شاخص‌های خطای R و MSE استخراج می‌گردد.

نتایج و بحث

مدل رگرسیون خطی: درگام اول مدل‌سازی رگرسیون خطی، انتخاب تابع برازش مناسب از بین توابع معروف شامل تابع خطی، درجه دو، درجه دو اصلاح شده و متقابل براساس حداکثر مقدار شاخص CMR صورت گرفته است. بهترین تابع، تابع درجه دو اصلاح شده با بیش‌ترین مقدار شاخص CMR می‌باشد. به‌منظور انتخاب نوع تابع وزن در روش مربعات وزنی دوگانه، مدل رگرسیون خطی این پژوهش را برای توابع وزن متفاوت اجرا و سپس مقادیر شاخص خطای CMR محاسبه گردید. تابع وزن Cauchy به‌دلیل دارا بودن بیش‌ترین مقدار شاخص CMR، به‌عنوان بهترین تابع وزن مدل رگرسیون خطی این پژوهش انتخاب گردید.

مدل رگرسیون خطی تعمیم یافته: با توجه به مقادیر نرخ شکست لوله داده‌های ثبت شده شبکه توزیع آب این پژوهش، تابع توزیع حاکم بر متغیر مستقل، تابع توزیع پواسون و نوع تابع اتصال مناسب برای تابع توزیع پواسون، تابع Log می‌باشد. با توجه به تابع توزیع پواسون و تابع اتصال لوگ، مدل خطی تعمیم‌یافته نرخ شکست لوله شبکه توزیع آب محدوده مطالعاتی این پژوهش تهیه‌گردید.

مدل رگرسیون بردار پشتیبان: به‌دلیل زیادبودن پارامترهای مدل، انتخاب بهترین مقادیر برای پارامترهای فوق فقط با استفاده از روش بهینه‌سازی ممکن می‌باشد. در این پژوهش همان‌طوری‌که در شکل ۳ نشان داده شده است بهترین روش بهینه‌سازی، روش LIQP همراه با تابع کرنل گوسی می‌باشد. مقادیر بهینه سایر پارامترهای مدل SVR این پژوهش در جدول ۲ ارائه شده است.



شکل ۳- تغییرمقدار شاخص CMR در روش‌های مختلف بهینه‌سازی مدل SVR با درنظر گرفتن توابع کرنل.

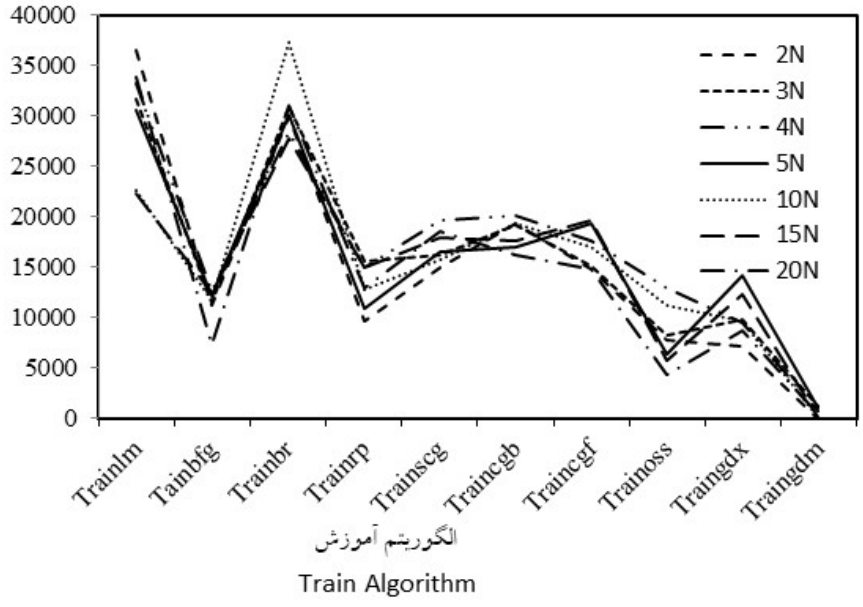
Figure 3. Change of CMR error index for different SVR optimization methods based on Kernel functions.

جدول ۲- مقادیر بهینه پارامترهای مدل SVR پژوهش حاضر.

Table 2. Optimal values of SVR model in the study.

روش بهینه‌سازی (Optimization method)	LIQP
تابع کرنل (Kernel function)	Gaussian
مقیاس کرنل (Kernel Scale)	0.0023
استاندارد (Standard)	Yes
Epsilon (σ)	0.005
Bias (b)	0.014
Box constrain (C)	309.98

شبکه عصبی مصنوعی پیش خور: به‌منظور انتخاب الگوریتم آموزش مناسب مسأله این پژوهش، ابتدا شبکه عصبی مصنوعی پیش‌خور با یک لایه مخفی، تعداد نرون مختلف و ده الگوریتم آموزشی درمحیط MATLAB-R2018b ساخته شد. سپس این مدل ۱۰۰ بار اجرا گردید و شاخص‌های MSE و R و در نتیجه CMR مربوط به داده‌های آموزشی و آزمایشی برای شبکه‌ها با تعداد نرون‌ها و الگوریتم‌های آموزش مختلف محاسبه گردید. همان‌طوری‌که در شکل ٤ نشان داده شده است، الگوریتم آموزش Trainlm به‌دلیل دارا بودن بیش‌ترین مقدار CMR دارای عملکرد مناسب‌تری بوده و بنابراین به‌عنوان الگوریتم آموزش شبکه عصبی مصنوعی پیش‌خور این پژوهش انتخاب گردید.

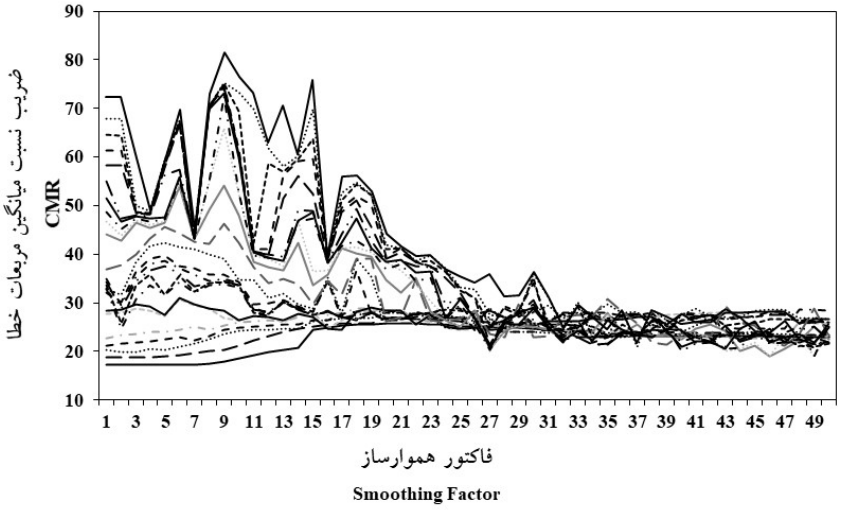


شکل ٤- مقایسه شاخص CMR برای الگوریتم‌های آموزش و تعداد نرون مختلف شبکه عصبی مصنوعی پیش‌خور.

Figure 4. Comparing the CMR error index of different training algorithm and neurons number of FFNN.

به‌منظور انتخاب مناسب‌ترین تعداد نرون‌های در لایه مخفی، با توجه به نوع تابع محرک و الگوریتم آموزشی انتخابی، شبکه‌های عصبی مصنوعی پیش‌خور تهیه‌شده از مرحله قبل، به‌ازای تعداد نرون‌های مختلف اجرا و مقدار شاخص خطای CMR آن‌ها محاسبه گردید. بهترین تعداد نرون‌های مربوط به شبکه عصبی مصنوعی این پژوهش، برابر ۱۰ بوده که مربوط به بیش‌ترین مقدار شاخص CMR می‌باشد؛ بنابراین مشخصات شبکه پیش‌خور این پژوهش شامل یک لایه ورودی با ۵ متغیر، یک لایه مخفی با ۱۰ نرون و تابع محرک Tansig، لایه خروجی با یک متغیر وابسته (نرخ شکست هر لوله) و تابع محرک خطی Purelin و الگوریتم آموزشی Trainlm می‌باشد.

شبکه عصبی مصنوعی پایه شعاعی: در این پژوهش برای تعیین بهترین فاکتور هموارساز، مدل شبکه عصبی مصنوعی RBF به‌ازای مقادیر مختلف فاکتور هموارساز بین ۱ و ۱۰۰ و هم‌چنین تعداد نرون‌های مختلف در محیط MATLABR2018b اجرا شده و سپس مقادیر شاخص CMR به‌ازای مقادیر مختلف فاکتور هموارساز و تعداد نرون لایه مخفی محاسبه شد. همان‌طور که در شکل ۵ نشان داده شده است مقادیر فاکتور هموارساز در محدوده ۱۷-٦ دارای بیش‌ترین مقدار شاخص CMR می‌باشند. به‌ازای فاکتورهای هموارساز بین ٦ و ۱۷، حد بالای تعداد نرون لایه مخفی برای آموزش مدل این پژوهش برابر ۹ به‌دست آمده است.



شکل ۵- روند تغییرات شاخص CMR در مقابل تغییرات فاکتور هموارساز به‌ازای تعداد نرون‌های مختلف در لایه مخفی شبکه RBF.

Figure 5. Changes in CMR index versus smoothing factor for different number of neurons in the hidden layer RBF network.

انتخاب مدل برتر: مقادیر شاخص‌های MSE، R، NS، MAE برای هریک از مدل‌ها برای داده‌های آموزش، آزمایشی و صحت‌سنجی در جدول‌های ۳ تا ۵ ارائه شده است.

برای انتخاب مدل پیش‌بینی برتر، شاخص‌های خطای مربوط به داده‌های صحت‌سنجی در مدل‌های مختلف با یکدیگر مقایسه می‌شوند. مقادیر شاخص خطای MSE داده‌های صحت‌سنجی مدل‌های مختلف نشان می‌دهد که مدل FFNN دارای کم‌ترین مقدار

خطای MSE می‌باشد که نشان‌دهنده دقت بیش‌تر این مدل نسبت به سایر مدل‌ها می‌باشد. علاوه برآن، مقدار شاخص R که نشان‌دهنده نکویی برازش مدل می‌باشد، در مدل FFNN برابر ۰/۶۹ بوده که در مقایسه با مدل‌های دیگر، بیش‌ترین مقدار می‌باشد. مقدار شاخص خطای NS=۰/٤٤ مدل FFNN نشان‌دهنده توانایی پیش‌بینی بیش‌تر مدل در مقایسه با سایر مدل‌ها می‌باشد؛ بنابراین از بین مدل‌های آماری و هوشمند این پژوهش، مدل FFNN به‌عنوان بهترین مدل جهت پیش‌بینی نرخ شکست لوله‌های شبکه انتخاب گردید.

جدول ۳- شاخص‌های خطای مدل‌های آماری و هوشمند مختلف برای داده‌های آموزش.

Table 3. Error Indexes of different soft and statistical model for train data.

MAE	NS	R	MSE	نام مدل/شاخص خطا Model Name/Error Index
0.0045	-1.91	0.49	0.0907	LR (رگرسیون خطی)
0.0044	-1.32	0.59	0.0868	GLR (رگرسیون خطی تعمیم یافته)
0.0036	0.13	0.55	0.0921	SVR (رگرسیون بردار پشتیبان)
0.202	0.46	0.68	0.0826	FFNN (شبکه عصبی مصنوعی پیش‌خور)
0.0024	0.788	0.91	0.0624	RBFNN (شبکه عصبی مصنوعی پایه‌شعاعی)

جدول ٤- مقایسه شاخص‌های خطای مدل‌های هوشمند مختلف برای داده‌های آزمایش.

Table 4. Error Indexes of different soft and statistical model for test data.

MAE	NS	R	MSE	نام مدل/شاخص خطا Error Index/Model Name
0.0027	0.493	0.71	0.059	LR (رگرسیون خطی)
0.0024	0.407	0.75	0.056	GLR (رگرسیون خطی تعمیم یافته)
0.003	0.58	0.93	0.067	SVR (رگرسیون بردار پشتیبان)
0.031	0.417	0.84	0.055	FFNN (شبکه عصبی مصنوعی پیش‌خور)
0.0045	0.23	0.49	0.077	RBFNN (شبکه عصبی مصنوعی پایه‌شعاعی)

جدول ۵- مقایسه شاخص‌های خطای مدل‌های هوشمند مختلف در مرحله صحت‌سنجی.

Table 5. Error Indexes of different soft and statistical model for verification data.

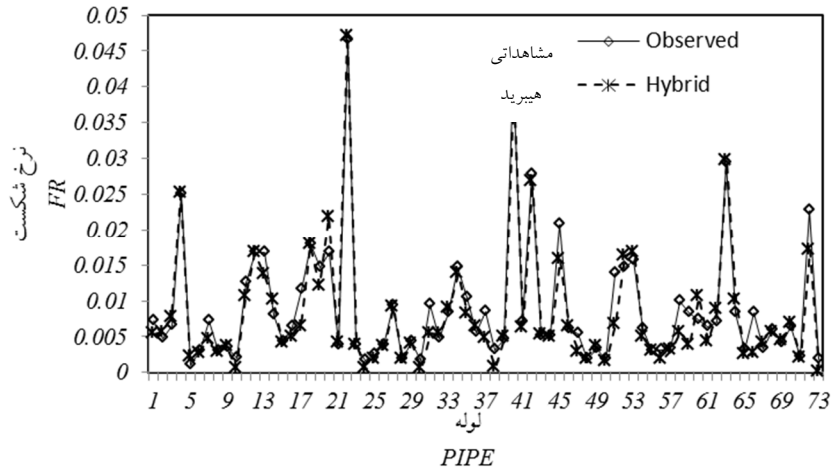
MAE	NS	R	MSE	نام مدل/شاخص خطا Error Index/Model Name
0.005	-1.77	0.51	0.0863	LR (رگرسیون خطی)
0.0057	-.353	0.35	0.0897	GLR (رگرسیون خطی تعمیم یافته)
0.0026	0.43	0.62	0.066	SVR (رگرسیون بردارپشتیبان)
0.0453	0.44	0.69	0.062	FFNN (شبکه عصبی مصنوعی پیش‌خور)
0.0051	-12.67	0.27	0.0931	RBFNN (شبکه عصبی مصنوعی پایه‌شعاعی)

مدل پیش‌بینی هیبرید: براساس مقایسه شاخص‌های MSE و R مربوط به مدل‌های مختلف مورد استفاده در این پژوهش، مدل FFNN به‌دلیل دارا بودن کم‌ترین مقدار MSE و بیش‌ترین مقدار R به‌عنوان بهترین مدل انتخاب شده‌است. در نهایت، نرخ شکست نهایی لوله‌ها با استفاده از مقادیر پیش‌بینی اولیه نرخ شکست لوله‌ها توسط هریک از مدل‌های آماری و هوشمند به‌عنوان مجموعه داده‌های متغیرهای مستقل ورودی و مقادیر نرخ شکست داده‌های مشاهداتی اولیه به‌عنوان متغیر وابسته خروجی و مدل شبکه عصبی مصنوعی FFNN پیش‌بینی گردید. مقادیر شاخص‌های آماری داده‌های آموزشی، آزمایشی و صحت‌سنجی مدل پیش‌بینی هیبرید در جدول ٦ ارائه شده است. مقایسه مقادیر مشاهداتی و پیش‌بینی‌شده نرخ شکت لوله‌های شبکه توزیع آب در مرحله صحت‌سنجی با استفاده از مدل هیبرید در شکل ٦ نشان داده شده است.

جدول ٦- شاخص‌های خطای مدل پیش‌بینی هیبرید برای داده‌های آموزش، آزمایش و صحت‌سنجی.

Table 6. Error Indexes of Hybrid predicting model for train, test and verification data.

MAE	NS	R	MSE	داده/شاخص خطا Error Index/Model Name
0.0262	0.99	0.99	0.0274	Train (آموزش)
0.0254	0.71	0.847	0.0576	Test (آزمایش)
0.0184	0.92	0.96	0.0462	Verification (صحت‌سنجی)



شکل ٦- مقادیر نرخ شکست مشاهداتی و پیش‌بینی شده توسط مدل هیبرید.

Figure 6. Observed and predicted failure rate values by Hybrid model.

نتیجه‌گیری کلی

به‌منظور پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب پژوهش حاضر، پنج مدل مختلف شامل سه مدل آماری (رگرسیون خطی، رگرسیون خطی تعمیم‌یافته، رگرسیون بردار پشتیبان) و دو مدل هوشمند (شامل شبکه عصبی مصنوعی پیش‌خور و شبکه عصبی مصنوعی پایه شعاعی) مورد بررسی قرار گرفتند. به‌منظور بررسی و مقایسه دقت، عملکرد و توانایی مدل‌های مختلف جهت پیش‌بینی نرخ شکست لوله‌های شبکه توزیع آب، در این پژوهش، شاخص‌ها و ابزارهای مختلفی استفاده شد. این شاخص‌ها و ابزارها از دو دیدگاه نکویی برازش و خطای میانگین به بررسی دقت و توانایی مدل‌ها می‌پردازند. علی‌رغم بیش‌تر بودن مقدار شاخص نکویی برازش (R) و کم‌تر بودن مقدار شاخص خطای میانگین (MSE) مدل شبکه عصبی مصنوعی پیش‌خور (خور) نسبت به سایر مدل‌های مورد بررسی در این پژوهش، این مدل به‌دلیل پایین‌بودن مقدار شاخص $R =$ ۰/۶۹ و بالابودن مقدار شاخص شاخص $SE =$ ۰/۶۲ دارای دقت و عملکرد مورد قبول جهت پیش‌بینی نرخ شکست لوله‌های شبکه مورد مطالعه نمی‌باشد. مدل هیبرید توسعه‌داده‌شده براساس رویکرد جدید ارائه شده در این پژوهش، با دارا بودن مقدار شاخص‌های R=۰/۹۶ و MSE=۰/۰۴۶۲، باعث افزایش شاخص خطای R به مقدار ۳۹ درصد و کاهش شاخص MSE به مقدار ۲۵ درصد در مقایسه با مدل شبکه عصبی مصنوعی پیش‌خور شده است که نشان‌دهنده عملکرد و دقت بهتر مدل هیبرید نسبت به سایر مدل‌ها می‌باشد. بنابراین، با توجه به عملکرد و توانایی بالای مدل هیبرید توسعه داده در این پژوهش، این مدل به‌عنوان مدل برتر پیش‌بینی انتخاب شده است و می‌تواند نرخ شکست لوله‌های شبکه توزیع آب شهری محدوده مطالعاتی پژوهش حاضر را با دقت بسیار بالایی پیش‌بینی نماید. هم‌چنین رویکرد پیشنهادی قابلیت ارائه در سایر شبکه‌های توزیع آب شهری را دارا می‌باشد.

تقدیر و تشکر

بدین‌وسیله از شرکت آب و فاضلاب استان گلستان به‌خاطر کمک و حمایت در جمع‌آوری اطلاعات و آمار موردنیاز این مقاله کمال تشکر و قدردانی را داریم.

داده‌ها و اطلاعات

منبع کلیه آمار و اطلاعات مربوط به شبکه توزیع آب مورد مطالعه در این مقاله، شرکت آب و فاضلاب استان گلستان می‌باشد.

تعارض منافع

در این مقاله تعارض منافعی وجود ندارد و این مسأله مورد تأیید همه نویسندگان است.

منابع

1. Aydogdu, M., and Firat. M. 2015. Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM methods. Water resources management 29: 5. 1575-1590.
2. Asnaashari, A., McBean, E. A., Gharabaghi, B., and Tutt, D. 2013. Forecasting watermain failure using artificial neural network modelling. Canadian Water Resour. J. 38: 1. 24-33.
3. Berardi, L., Giustolisi, O., Kapelan, Z., and Savic, D.A. 2008. Development of pipe deterioration models for water distribution systems using EPR. J. Hydroinf. 10: 2. 113-126.

4.Farmani, R., Kakoudakis, K., Behzadian Moghadam, K., and Butler, D. 2017. Pipe failure prediction in water distribution systems considering static and dynamic factors. Procedia Engineering, 186: 117-126.
5.Fares, H., and Zayed, T. 2010. Hierarchical fuzzy expert system for risk of failure of water mains. J. Pipeline Syst. Engin. Prac. 1: 1. 53-62.
6.Faris Hamdala, K., and Sagar, G. 2016. Statistical Analysis of Pipe Breaks in Water Distribution Systems in Ethiopia, the Case of Hawassa.
7.Ghasemnejad, S., and Nodoshan, M.S. 2014. Prediction of water distribution network pipes Vulnerability by developing statistical models. Iranian Congress of Water and Wastewater Engineering. (In Persian)
8.Ho, C.I., Lin, M.D., and Lo, S.L. 2010. Use of a GIS-based hybrid artificial neural network to prioritize the order of pipe replacement in a water distribution network. Environmental monitoring and assessment, 166: 1-4. 177-189.
9.Islam, M.S., Sadiq, R., Rodriguez, M.J., Francisque, A., Najjaran, H., and Hoorfar, M. 2011. Leakage detection and location in water distribution systems using a fuzzy-based methodology. Urban Water J. 8: 6. 351-365.
10.Jafar, R., Shahrour, I., and Juran, I. 2010. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. Mathematical and Computer Modelling, 51: 9-1. 1170-1180.
11.Kerwin, S., de Soto, B.G., and Adey, B.T. 2019. Performance comparison for pipe failure prediction using artificial neural networks. In 6th International Symposium on Life-Cycle Civil Engineering, IALCCE 2018 (pp. 1337-1342). CRC Press/Balkema.
12.Kakoudakis, K., Behzadian, K., Farmani, R., and Butler, D. 2017. Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. Urban Water J. 14: 7. 737-742.
13.Kleiner, Y., and Rajani, B. 2002. Forecasting variations and trends in water-main breaks. J. Infrastruc. Syst. 8: 4. 122-131.
14.Kapelan, Z.S., Savic, D.A., and Walters, G.A. 2003. A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks. J. Hydr. Res. 41: 5. 481-492.
15.Motiee, H., and Ghasemnejad, S. 2019. Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. Water Supply, 19: 3. 695-702.
16.Mounce, S.R., Day, A.J., Wood, A.S., Khan, A., Widdop, P.D., and Machell, J. 2002. A neural network approach to burst detection. Water science and technology, 45: 4-5. 237-246.
17.Pandey, P., Dongre, S., and Gupta, R .2020. Probabilistic and fuzzy approaches for uncertainty consideration in water distribution networks–a review. Water supply. 20: 1. 13-27.
18.Robles-Velasco, Alicia, Pablo Cortés, Jesús Muñuzuri, and Luis Onieva. "Prediction of pipe failures in water supply networks using logistic regression and support vector classification." Reliability Engineering & System Safety. 196 (2020): 106754.
19.Rohan Dipakkumar Kania, S.S. 2017. Risk assessment of water distribution network using fuzzy based system. Inter. J. Adv. Res. Engin. Sci. Manage.
20.Rajani, B., and Kleiner, Y. 2001. Comprehensive review of structural deterioration of water mains: physically based models. Urban water, 3: 3. 151-164.
21.Rogers, P.D., and Grigg, N.S. 2009. Failure assessment modeling to prioritize water pipe renewal: two case studies. J. Infrastruc. Syst. 15: 3. 162-171.
22.Sadiq, R., Kleiner, Y., and Rajani, B. 2007. Water quality failures in distribution networks-risk analysis using fuzzy logic and evidential reasoning. Risk Analysis: An Inter. J. 27: 5. 1381-1394.
23.Shamir, U., and Howard, C.D. 1979. An analytic approach to scheduling pipe replacement. J. Amer. Water Work Assoc. 71: 5. 248-258.

24. Shirzad, A., Tabesh, M., and Farmani, R. 2014. A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. KSCE J. Civil Engin. 18: 4. 941-948.
25. Soltani, J., and Tabari Rezapour, M. 2012. Determination of effective parameters in pipe failure rate in water distribution system using the combination of artificial neural networks and genetic algorithm. J. Water Waste 83: 2-18.
26. Sattar, A.M., Gharabaghi, B., and McBean, E.A. 2016. Prediction of timing of watermain failure using gene expression models. Water resources management, 30: 5. 1635-1651.
27. Soltanjalili, M., Bozorg-Haddad, O., and Mariño, M.A. 2011. Effect of breakage level one in design of water distribution networks. Water resources management, 25: 1. 311-337.
28. Sacluti, F.R. 1999. Modelling water distribution pipe failures using artificial neural networks.
29. Tabari, M.M.R., and Malekpour Shahraki, M.M. 2018.Reservoir Water Level Prediction Using Supervised Intelligent Committee Machine Method, Case Study: Karaj Amirkabir Dam. Iran-Water Resources Research, 14: 5. 15-30 (In Persian)
30. Tavakoli, R., Najafi, M., and Sharifara, A. 2019. Artificial Neural Networks and Adaptive Neuro-fuzzy Models for Prediction of Remaining Useful Life. arXiv preprint arXiv:1909.02115.
31. Tu, M.Y., Tsai, F.T.C., and Yeh, W.W.G. 2005. Optimization of water distribution and water quality by hybrid genetic algorithm. J. Water Resour. Plan. Manage. 131: 6. 431-440.
32. Wang, Y., Zayed, T., and Moselhi, O. 2009. Prediction models for annual break rates of water mains. J. Perform. Construc. Facilit. 23: 1. 47-54.
33. Xu, Q., Chen, Q., Ma, J., and Blanckaert, K. 2014. Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. J. Hydro-Environ. Res. 7: 2. 134-140.
34. Xu, Q., Chen, Q., Li, W., and Ma, J. 2011. Pipe break prediction based on evolutionary data-driven methods with brief recorded data. Reliability Engineering & System Safety, 96: 8. 942-948.

Gorgan University of Agricultural Sciences and Natural Resources

Research Full Paper

J. of Water and Soil Conservation, Vol. 27(5), 2021
http://jwsc.gau.ac.ir
DOI: 10.22069/jwsc.2021.17670.3319

New Approach for Prediction of Water Distribution Network Pipes Failure Based on a Intelligent Hybrid Model (Case Study: Gorgan Water Distribution Network)

S.M. Jafari[1], *A. Zahiri[2], O. Bozorg Hadad[3] and M. Mohammad Rezapour Tabari[4]

[1]Ph.D. Student, Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Iran,
[2]Associate Prof., Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Iran,
[3]Professor, Dept. of Water Engineering, University of Tehran, Iran,
[4]Associate Prof., Dept. of Civil Engineering, Faculty of Technical and Engineering, University of Mazandaran, Mazandaran, Iran
Received: 02.09.2020; Accepted: 08.26.2020

Abstract

Background and Objectives: Urban water distribution networks consider as one of the essential infrastructural facilities and equipment in urban areas. The pipes are one of the primary and essential components of a water distribution network break during operation due to various factors. So, developing models for pipes failure rate prediction can be one of the most crucial tools for managers and stakeholders to the optimal operation of the water distribution network. In the last decade, various studies have performed to predict the failure rate of water distribution pipes using statistical and soft models - each of which has strengths and weaknesses. This study aims to present a new approach based on the development of a hybrid prediction model, considering the capabilities of soft and statistical models, to more accurately predict the water distribution network pipes failure rate compared to statistical and soft models used in previous research.

Materials and Methods: In order to achieve the study goals, 4-year (2015-2018) time duration statistics of Gorgan water distribution network characteristics including diameter, length, age, depth of installation, and the number of pipe failures used to predict future pipes failure rates. To predict the failure rate of water distribution network pipes, five different models, including three statistical models (linear regression, generalized linear regression, support vector regression) and two soft models (feed-forward neural network, and radial basis function neural network) has studied. Optimal parameters of the models were selected based on appropriate statistical error indicators, including correlation coefficient (R), Mean Square Error (MSE), and Correlation Mean square error Ratio (CMR) for the training and testing data. In order to select the best model from different models to predict the failure rate of network pipes, the values of R and MSE indicators of the above models were calculated in the validation stage and compared with each other. Finally, to predict pipes failure rate more accurately, a new approach is developed based on the hybrid prediction model in which the predicted values of pipe failure rates by statistical and soft computing models considered as independent variables of the best model inputs and the observed values of failure rates as dependent variables of the best model outputs.

* Corresponding Author; Email: zahiri.areza@gmail.com

Results: Comparing the values of R and MSE indicators of each statistical and soft computing model used in this study in the validation phase show that these models cannot predict the pipes' failure rate with reasonable accuracy. Feedforward neural network model with the highest R = 0.69 and the lowest MSE = 0.062 values has the best estimates. Using the new approach developed based on hybrid soft and statistical models, the R index is equal to 0.96, and the MSE index is equal to 0.046.

Conclusion: A significant increase in the R index (39%) and decrease in the MSE index (25%) through using the proposed hybrid approach compared to the feed-forward neural network model demonstrates that using the new approach provides perfect accuracy prediction of the pipes failure rate of the water distribution network.

***Keywords*:** Gorgan, Hybrid Model, Intelligent Model, Pipe Failure Rate, Urban Water Distribution Network