

مدل سازی مواد جامد محلول با استفاده از رویکرد مبتنی بر هوش مصنوعی بر پایه الگوریتم های پیش پردازش کننده

سعید پی پل زاده^۱، رضا مستوری^{۲*} و نازنین شاه کرمی^۳

چکیده

کیفیت آب یکی از مهم ترین عوامل موثر در زندگی سالم و حیات بشر است. از این رو، شناسایی مواد جامد محلول (TDS) یکی از مهم ترین عوامل آن می باشد؛ که بسیاری از برنامه های توسعه منابع آب در شناسایی این عوامل اجرا خواهند شد. پیش بینی دقیق پارامترهای کیفیت آب یک نیاز اساسی برای مدیریت کیفیت آب، سلامت انسان، مصرف عمومی و مصارف خانگی است. در این مطالعه، از یک الگوریتم پیش پردازش داده جدید، EEMD، برای تخمین یکی از پارامترهای مهم کیفیت آب، بنام TDS استفاده شد. پذیرش و قابلیت اطمینان مدل های پیشنهادی (به عنوان مثال، شبکه های عصبی مصنوعی (ANN)، EEMD-ANN، ماشین بردار پشتیبان (SVM) و (EEMD-SVM) با استفاده از پنج معیار عملکرد و نمودارهای تصویری ارزیابی شد. مقایسه نتایج مدل های مستقل و ترکیبی نشان داد که الگوریتم های پیش پردازنده می توانند عملکرد مدل SVM ترکیبی را برای برآورد پارامتر کیفی TDS افزایش دهد. به عنوان مثال، مدل EEMD-SVM با RMSE برابر ۲۰/۲۳ برای مرحله آموزش و RMSE برابر ۲۷/۲۹ برای مرحله آزمایش در ایستگاه ورنه و RMSE=۴۵/۲۶ برای مرحله آموزش و RMSE=۴۰/۰۶ برای مرحله آزمایش در ایستگاه گرم رود، از سایر مدل های ترکیبی و مستقل عملکرد بهتری داشت. از این رو، می توان روش ترکیبی ماشین های بردار پشتیبان بر پایه الگوریتم پیش پردازش کننده EEMD را به عنوان یک مدل برتر به تصمیم گیران جهت برنامه ریزی و مدیریت در زمینه تشخیص کیفیت آب رودخانه ها پیشنهاد کرد.

واژه های کلیدی: الگوریتم های پیش پردازش داده ها، پارامترهای کیفیت آب، جامدات جامد محلول، شبکه های عصبی مصنوعی، ماشین بردار پشتیبان.

ارجاع: پی پل زاده س.، مستوری ر. و شاه کرمی ن. ۱۴۰۱. مدل سازی مواد جامد محلول با استفاده از رویکرد مبتنی بر هوش مصنوعی بر پایه الگوریتم های پیش پردازش کننده. مجله پژوهش آب ایران. ۳۷-۲۵: ۴۴. DOI: <https://dx.doi.org/10.22034/iwrj.2022.11217>

۱- دانشجوی دکترا مهندسی و مدیریت منابع آب، دانشکده مهندسی عمران، دانشگاه آزاد اسلامی واحد اراک.

۲- استادیار گروه مهندسی منابع آب، دانشگاه آزاد اسلامی واحد اراک.

۳- استادیار گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه اراک.

*نویسنده مسئول: r-mastouri@iau-arak.ac.ir

تاریخ پذیرش: ۱۴۰۰/۰۹/۲۰

تاریخ دریافت: ۱۴۰۰/۰۴/۱۲

مقدمه

رودخانه ها مهم ترین منابع حیاتی طبیعت هستند. امروزه در دنیا، آب و منابع آبی یکی از پایه های اصلی توسعه پایدار به شمار می روند. در مواردی که رودخانه ها به عنوان یکی از منابع تأمین کننده نیاز جوامع بشری مطرح می شوند، علاوه بر کمیت و میزان آورد رودخانه، کیفیت آب نیز جزو پارامترهای مهم مورد توجه قرار می گیرد (میثاقی و محمدی، ۱۳۸۳). همچنین کشور ایران در زمره کشورهای نیمه خشک دنیا و جزو چند کشور دارای سرانه پایین آب است؛ به طوری که مطابق با پیش بینی های بین المللی، در دهه ۱۱، از نظر سرانه آب، به مرز تنش آبی و تا سال ۱۴۰۰ به حد بحران آب خواهیم رسید (حسنی و همکاران، ۱۳۹۵). استان مازندران به دلیل دارا بودن شرایط آبی مناسب، از قطب های مهم کشاورزی محسوب می شود که در صورت بی توجهی به حفظ این شرایط، پایداری بخش کشاورزی این استان با خطر مواجه می شود. تجن یکی از رودخانه های استان مازندران است که علاوه بر تأمین آب شبکه های آبیاری و زهکشی دشت های اطراف، نقش اساسی در تأمین آب شرب شهر ساری و روستاها و شهرهای اطراف آن ایفا می کند. کیفیت آب تجن به دلایل مختلف از بالادست به پایین دست در حال کاهش است؛ بنابراین بررسی این رودخانه به عنوان یکی از رودخانه های مهم استان، از اولویت خاصی برخوردار است (یوسفی، ۱۳۹۴؛ نوشادی و همکاران، ۱۳۸۶).

در دو دهه اخیر، با ظهور روش های مبتنی بر هوش مصنوعی نظیر شبکه های عصبی مصنوعی، الگوریتم ژنتیک و منطق فازی کاربرد مدل های جعبه سیاه به منظور مدل سازی و پیش بینی پارامترهای مختلف در مهندسی آب به شدت توسعه یافته است (ژیانگ و همکاران، ۲۰۲۰)؛ به طوری که در مطالعات مربوط به پیش بینی پارامترهای مختلف منابع آب مورد استفاده قرار گرفته و محققان مختلف بر دقت بالای این روش در مقایسه با سایر مدل ها تأکید کرده اند (رضایی-بالی و همکاران، ۲۰۲۰؛ میر و دندی، ۱۹۹۶). به طور مثال در ایران، کرمی و همکاران (۱۳۸۵)، با استفاده از پارامترهای اندازه گیری شده در ایستگاه های پل شالو، ملاتانی رامین، اهواز، فارسیات و به کارگیری شبکه های عصبی مصنوعی کیفیت آب رودخانه کارون را پیش بینی کردند. عظیمی (۱۳۸۷)، به شناخت و ارزیابی کیفیت آب رودخانه جاجرود پرداخته است. با

نتایج به دست آمده از این تحقیق، اثر تصفیه خانه ها و زهکش ها بر تغییرات اکسیژن محلول (DO^1) و کل جامدات محلول (TDS) در طول مسیر بررسی شده و اولویت های اجرایی مدیریت جامع منابع آب حوزه آبریز بدست آمد. همچنین امیری و امیری (۱۳۹۴)، با استفاده از روش شبکه عصبی مصنوعی به پیش بینی پارامترهای کیفی رودخانه پرداختند. طی بررسی های انجام شده در این تحقیق، مشاهده شد پیش بینی پارامترهای کیفی جریان آب رودخانه به ازای میزان بازدهی شبکه برای پارامترهای TDS، SAR و EC به ترتیب ۰/۹۶، ۰/۹۳ و ۰/۹۹ است که بیانگر دقت بالای شبکه عصبی مصنوعی در پیش بینی پارامترهای کیفی آب رودخانه است. رجایی و رحیمی (۱۳۹۵)، عملکرد مدل های شبکه عصبی مصنوعی (ANN^2)، مدل ترکیبی شبکه عصبی-موجک (WANN) و رگرسیون خطی چندمتغیره (MLR^3)، در پیش بینی یک ماه آینده یون نیتрат و کلراید ایستگاه ورودی آبگیر بیلقان واقع در رودخانه کرج، مورد ارزیابی قرار دادند. نتایج حاکی از دقت بالای مدل ترکیبی شبکه عصبی-موجکی نسبت به دو مدل دیگر بوده است.

در خارج از ایران، پالانی و همکاران (۲۰۰۰)، به پیش بینی مشخصات کیفی شامل شوری، دما، اکسیژن محلول و کلروفیل در آب های ساحلی با استفاده از شبکه های عصبی مصنوعی پرسپترون چندلایه پرداختند. این تحقیق توانایی بالای مدل شبکه عصبی در شبیه سازی پارامترهای کیفی آب را نشان داد. سینگ و همکاران (۲۰۰۹) از مدل شبکه عصبی مصنوعی برای تخمین غلظت DO و اکسیژن خواهی بیوشیمیایی در رودخانه ای در هند استفاده کردند. ورودی مدل ارائه شده عبارت بود از نتایج اندازه گیری ۱۱ پارامتر کیفی آب در طی ۱۰ سال به طور ماهانه در ۸ ایستگاه مختلف. آن ها عملکرد مدل شان را با استفاده از ضریب تشخیص، خطای جذر میانگین مربعات و مقایسه با مقادیر اندازه گیری شده ارزیابی کردند. نتایج مدل، تطبیق خوبی با مقادیر اندازه گیری شده و مورد انتظار برای غلظت های رودخانه را نشان داد. عمر فاروق (۲۰۱۰)، در تحقیقی از مدل ترکیبی شبکه عصبی مصنوعی و مدل ARIMA برای پیش بینی سری زمانی کیفیت آب استفاده کرد. نتایج این تحقیق حاکی از توانایی بالای مدل ترکیبی شبکه عصبی

1- Dissolved Oxygen

2- Artificial Neural Network

3- Multivariable Linear Regression

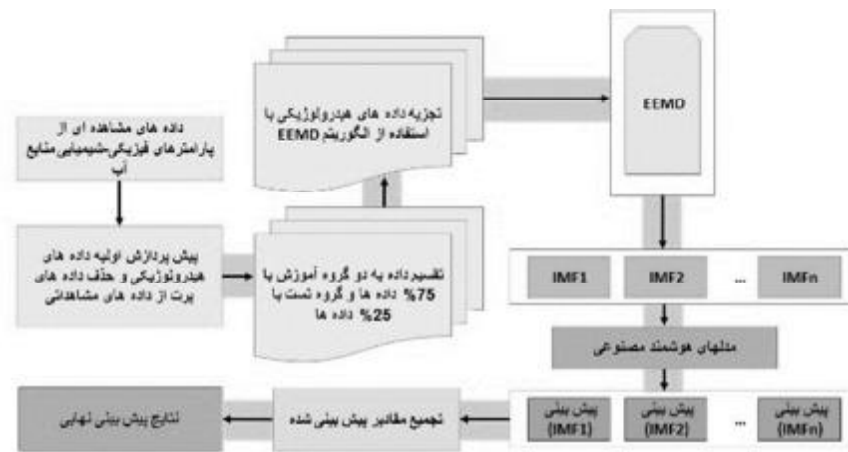
پارامترهای کیفی آب رودخانه مورد بررسی قرار گرفت. هدف اصلی این تحقیق، ارزیابی مقادیر برآوردشده با روش‌های نوین هوش مصنوعی برای تعیین رابطه یا روابط دقیق‌تر بین پارامترهای هیدرولوژیک و کیفی است تا براساس آن بتوان مقدار آلاینده‌گی آب رودخانه را با دقت بیشتری برآورد کرد.

مواد و روش‌ها

در این بخش به توضیح منطقه مورد مطالعه، یعنی حوضه رودخانه تجن پرداخته می‌شود و پس از آن، هریک از روش‌ها و الگوریتم‌های پیشنهادی در این مقاله به‌صورت مختصری ارائه می‌شوند. در این مقاله، دو روش هوشمند مصنوعی یعنی شبکه عصبی مصنوعی و ماشین‌های بردار پشتیبان به‌همراه یک الگوریتم پیش‌پردازش‌کننده داده‌های هیدرولوژیک به‌کار گرفته شد که توسعه گام‌به‌گام مدل‌های تلفیقی برای مدل‌سازی و پیش‌بینی مواد جامد محلول گام‌به‌گام در شکل ۱ نشان داده شده است.

مصنوعی و ARIMA به‌منظور پیش‌بینی سری‌های زمانی پیچیده و غیرخطی است. در تحقیقی، تومیک و همکاران (۲۰۱۸) در مطالعه‌ای به پیش‌بینی اکسیژن محلول (DO) با استفاده از یک نوع از روش‌های شبکه عصبی بهینه‌سازی‌شده در رودخانه دانوب پرداختند. نتایج به‌دست‌آمده نشان داد که روش پیشنهادی با ضریب تعیین بیشتر از ۰/۸۶ عملکرد مؤثری در پیش‌بینی اکسیژن محلول داشته است. شرافتی و همکاران (۲۰۲۰) با استفاده از روش‌های یادگیری ماشین گروهی به پیش‌بینی سه پارامتر TDS، BOD و COD در تصفیه‌خانه‌ای از استان قم پرداختند. در این تحقیق از سه روش ABR، GBR و RFR برای این پارامترهای کیفی آب در نظر گرفته شد. در تخمین پارامترهای کیفی آب، نتایج نشان داد که روش ABR با ضریب همبستگی ۰/۹۶، ۰/۹ و ۰/۷۵ به‌ترتیب برای BOD، TDS و COD بهترین مدل شناخته شد.

در این پژوهش، به‌صورت موردی رودخانه تجن و ایستگاه هیدرومتری واقع بر این رودخانه به‌منظور ارزیابی تخمین



شکل ۱- چارچوب مدل‌های تلفیقی هوشمند برای پیش‌بینی مواد جامد محلول در حوضه رودخانه تجن

منطقه مورد مطالعه

تجن از رودخانه‌های «مهم حوزه آبریز دریای مازندران است که از ارتفاع ۳۲۵۱ متری کوه‌های هزار جریب در دامنه شمالی سلسله جبال البرز در جنوب شهرستان ساری از چشمه‌های متعددی در دهستان پشتکوه سرچشمه می‌گیرد» که از شمال به دریای خزر «از شرق به حوضه رودخانه دارابکلا و نكاء رود» از جنوب به استان سمنان و از غرب به حوضه سیاه‌رود و رودخانه تالار محدود می‌شود. طول شاخه اصلی رودخانه، ۱۷۲ کیلومتر و سه

شاخه اصلی آن دودانگه و چهاردانگه و ظالم‌رود نام دارند که بر روی شاخه دودانگه، سد مخزنی شهید رجایی و سد خاکی فریم اجرا شده است و در مسیر شمالی تا رسیدن به دریای خزر از شهرستان ساری می‌گذرد (فراهی و همکاران، ۱۳۹۷). مساحت حوزه آبریز تا ایستگاه کردخیل واقع در پایاب رودخانه، حدود ۴۰۰۰ کیلومترمربع است که حدود پنجاه درصد آن مناطق کوهستانی پوشیده از جنگل است و رژیم جریان رودخانه، برفی-بارانی بوده که دارای جریان پایهای دائم است و میزان متوسط سالانه دبی

شبکه را محاسبه می کند و انواع متفاوتی دارد و می تواند حالتی خطی یا غیرخطی داشته باشد. تانژانت سیگموئید و لوگاریتمیک سیگموئید برای حالات غیرخطی و تابع تحریک پیورلین برای حالت خطی است که به ترتیب با معادلات (۲) تا (۴) تعیین می شوند (آرفا و همکاران، ۲۰۱۱؛ رضایی بلف و کیسی، ۲۰۱۸). شبکه های عصبی پیش خور، مشهورترین و پرکاربردترین شبکه عصبی مصنوعی هستند که با نام های متفاوتی، از جمله شبکه عصبی چندلایه پرسپترون شناخته شده اند. شبکه عصبی پرسپترون توسط فرانک روزنبلات در سال ۱۹۵۸ میلادی ابداع شد (آرفا و همکاران، ۲۰۱۱). این شبکه ها دارای سه لایه ورودی، پنهان و خروجی می شوند. در شبکه های عصبی پیش خور، سیگنال های ورودی از واحدهای ورودی به سمت واحدهای خروجی در جریان هستند (کایا، ۲۰۱۰؛ حسینی و همکاران، ۱۳۹۶).

$$a = \text{tansig}(n) = \frac{2}{(1 + \exp(-2 * n)) - 1} \quad (2)$$

$$a = \text{logsig}(n) = \frac{1}{(1 + \exp(-n))} \quad (3)$$

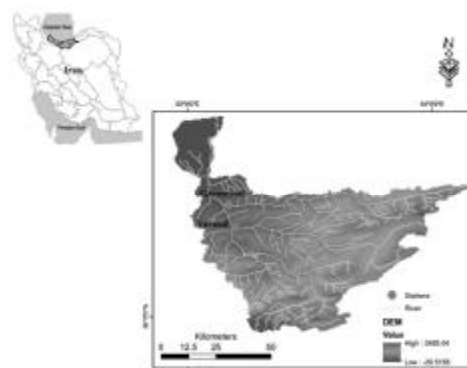
$$a = \text{purelin}(n) = n \quad (4)$$

روش طبقه بندی ماشین بردار پشتیبان (SVM)

روش SVM یکی از روش های یادگیری با نظارت است که برای طبقه بندی و رگرسیون استفاده می شود و بر این اساس عمل می کند که هیچ گونه اطلاعاتی از چگونگی توزیع داده ها وجود نداشته باشد. ویژگی اصلی این روش توانایی بالا در استفاده از نمونه های تعلیمی کمتر و رسیدن به دقت بالاتر در مقایسه با سایر روش های طبقه بندی پیشین است (فام و همکاران، ۲۰۱۶). روش طبقه بندی ماشین بردار پشتیبان که یکی از روش های طبقه بندی خطی است، بهترین ابرسطحی را پیدا می کند که با حداکثر فاصله ممکن، داده های مربوط به دو طبقه را از هم تفکیک کند.

در نحوه ساخت ابرسطح جداکننده توسط ماشین بردار پشتیبان، ابتدا یک پوسته در اطراف نقاط هرکدام از کلاس ها در نظر بگیرید. ایده اصلی این است که یک جداکننده مناسب انتخاب شود؛ منظور، جداکننده ای است که بیشترین فاصله را با نقاط همسایه از هر دو طبقه دارد. این بردارها، بردارهای پشتیبان نام دارند و فرمول ریاضی این دو ابرسطح موازی که مرز جداکننده را تشکیل می دهند، در معادلات (۵) و (۶) نشان داده شده است:

رودخانه ایستگاه کردخیل در دوره شاخص آماری ۱۳۴۹-۱۳۴۸ تا ۱۳۸۰-۱۳۷۹، ۱۴/۶ مترمکعب در ثانیه است که آورده سالانه آن ۵۶۰ میلیون مترمکعب است. متوسط بارندگی سالانه حوضه در دوره آماری ۴۰ ساله (منتهی سال ۱۳۷۹-۱۳۸۰) به میزان ۸۹۳ میلی متر است (پارساخو، ۱۳۹۱). همچنین، شکل ۲ نقشه توپوگرافی منطقه مورد مطالعه را نشان می دهد.



شکل ۲- توپوگرافی منطقه مورد مطالعه - حوضه رودخانه تجن شبکه های عصبی مصنوعی (ANN)

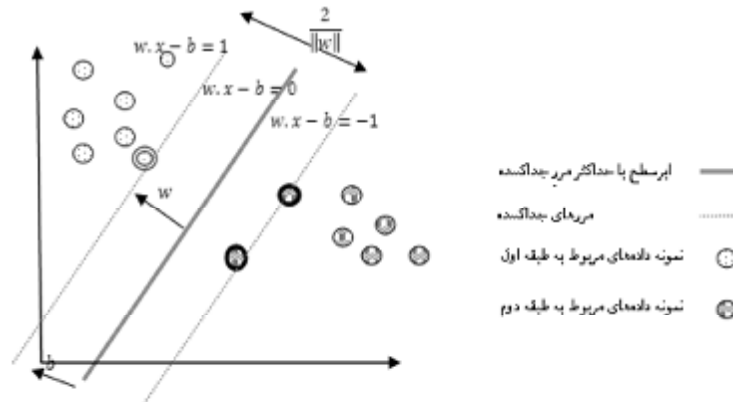
المان پردازشگر پایه شبکه های عصبی، نرون های مصنوعی یا گره نامیده می شوند. در مدل ریاضی ساده شده عصب، تأثیرات سیناپس ها با وزن های اتصال که اثر سیگنال های ورودی را بیان می کنند، ارائه می شوند. همچنین خاصیت غیرخطی یا خطی عصب ها به وسیله تابع انتقال (فعال سازی یا تحریک) نمایندگی می شود. عصب با دریافت سیگنال از لایه قبل از خود تحریک شده و جمع وزن های سیگنال های ورودی را محاسبه کرده و توسط تابع تحریک به لایه بعدی منتقل می کند. قابلیت آموزش شبکه عصبی با تنظیم وزن ها با توجه به الگوریتم آموزش به دست می آید (نهدی و همکاران، ۲۰۰۱). ورودی یک عصب از ضرب خروجی عصب های متصل به آن در وزن ها حاصل می شود. حاصل جمع وزن های ورودی با معادله (۱) محاسبه می شود.

$$(net)_j = \sum_{i=1}^n w_i o_i + b \quad (1)$$

که $(net)_j$ حاصل جمع وزن عصب j ام برای ورودی های لایه قبل است که دارای n عصب است. w_i وزن بین عصب j ام در لایه قبل و عصب i ام در لایه قبل است. b کمیت بایاس نام دارد و ثابتی برای تنظیمات داخلی عصب j ام است. تابع محرک، خروجی یک لایه از

$$w \cdot x - b = -1 \quad (۶)$$

$$w \cdot x - b = 1 \quad (۵)$$



شکل ۳- ابرسطح با حداکثر مرز جداکننده مرزهای جداکننده به همراه طبقه‌بندی نمونه داده‌های مربوط به دو طبقه متفاوت

است. نتایج تجزیه EEMD می‌تواند داده‌های مورد نیاز برای تکمیل تبدیل هیلبرت را فراهم کند. در موارد زیادی، EEMD با تبدیل هیلبرت ترکیب شده و به نام تبدیل هیلبرت-هوانگ (HHT) شناخته می‌شود. پس از مطالعه مفهوم فرکانس آنی در تبدیل هیلبرت، هوانگ مفهوم تازه‌ای از تابع مُد ذاتی (IMF) را که مؤلفه اساسی هر مجموعه داده پیچیده است، ارائه کرد (وو و هوانگ، ۲۰۰۹؛ رضایی-بالیف، ۲۰۱۹). این الگوریتم از یک روند غربال تکرارشونده استفاده می‌کند که با موفقیت می‌تواند میانگین محلی را از یک سیگنال کم کند. این روند غربال‌گری در ادامه بیان می‌شود (تانگ و همکاران، ۲۰۲۰):

1. اکستریمم محلی سیگنال (ماکزیمم، مینیمم) تعیین می‌شود.

2. مقدار ماکزیمم به یک تابع درونیابی متصل شده و پوش بالایی سیگنال را می‌سازد.

3. مقدار مینیمم به یک تابع درونیابی متصل شده و پوش پایینی سیگنال را می‌سازد.

4. میانگین محلی به عنوان نصف اختلاف بین پوش بالایی و پایینی محاسبه می‌شود.

5. میانگین محلی از سیگنال کم می‌شود.

6. این روند بر باقی‌مانده تکرار می‌شود.

روند غربال‌گری تا زمانی که سیگنال به تعریف یک IMF برسد، تکرار می‌شود. سپس IMF، از سیگنال اصلی، کسر شده و روند غربال بر باقی‌مانده تکرار می‌شود. این کار تا زمانی که باقی‌مانده، یک تابع یکنواخت شود، ادامه

باید از قرارگرفتن نقاط داده در ناحیه درون مرز جلوگیری کرد و برای این کار یک محدودیت ریاضی به تعریف فرمال اضافه می‌شود (اسچولکوف و مولا، ۲۰۰۲). برای هر i ، با اعمال محدودیت‌های زیر اطمینان حاصل می‌شود که هیچ نقطه‌ای در مرز قرار نمی‌گیرد:

برای داده‌های مربوط به طبقه اول:

$$w \cdot x_i - b \geq 1 \quad (۷)$$

برای داده‌های مربوط به طبقه دوم:

$$w \cdot x_i - b \leq -1 \quad (۸)$$

می‌توان این محدودیت را به صورت معادله زیر نشان داد:

$$c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \quad (۹)$$

با در نظر گرفتن محدودیت زیر و به حداقل رساندن w ، مسئله بهینه‌سازی بدین شکل تعریف می‌شود:

$$c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \quad (۱۰)$$

الگوریتم تجزیه مُد تجربی (EEMD)

تجزیه مُد تجربی (EEMD) یک روش تجزیه اطلاعات برای استخراج سیگنال‌ها در اطلاعات تولیدشده توسط عملیات غیرخطی و غیرایستا است که شامل جداسازی محلی و کامل برگرفته از اطلاعات از یک سیگنال در نوسان‌های سریع و آهسته است که توسط نوردن ای-هوانگ و همکاران در سال ۱۹۹۸، هنگامی که هوانگ مشغول کار در ناسا بود، ابداع شد. ابداع روش EEMD در نتیجه مطالعه تبدیل هیلبرت و بهبود یافته روش EMD (توسط نوردن ای-هوانگ و همکاران در سال ۱۹۹۸)

که O_i مقدارهای مشاهده ای، P_i مقدارهای پیش بینی، $\bar{O}$ میانگین داده های مشاهده ای، $\bar{P}$ میانگین داده های پیش بینی و M تعداد داده ها است.

نتایج و بحث

تجزیه و تحلیل پارامترهای فیزیکی-شیمیایی کیفیت آب

برای پیش بینی مقادیر TDS با استفاده از روش های هوشمند مصنوعی، مهم ترین و اساسی ترین گامی که مستقیماً بر عملکرد مدل ها تأثیر گذارند، انتخاب مناسب و بهینه پارامترهای ورودی به مدل است که هم از پیچیدگی بیهوده مدل برای پیش بینی مقادیر TDS و هم از اتلاف بیهوده زمان در اجرای برنامه جلوگیری شود. در این بخش به تجزیه و تحلیل پارامترهای ورودی و خروجی به مدل های هوشمند در جهت پیش بینی TDS در مقیاس ماهانه در دو ایستگاه ورنه و گرمورد پرداخته خواهد شد. براین اساس ماتریس ضریب همبستگی پیرسون^۵ پارامترهای فیزیکی-شیمیایی کیفیت آب برای هر دو ایستگاه ورنه و گرمورد در شکل ۴ نشان داده شده است.

برای تجزیه و تحلیل بیشتر مقادیر TDS، برخی از پارامترهای آماری برای کل مجموعه داده های دو ایستگاه ورنه و گرمورد نظیر انحراف معیار، چولگی، میانگین، ماکزیمم، مینیمم در جدول ۱ آمده است. براساس نتایج به دست آمده از ضریب پیرسون در دو ایستگاه گرمورد و ورنه، پارامترهای Na و Cl بیشترین تأثیر را بر روی TDS در مقیاس ماهانه داشته اند و از طرفی پارامتر Ca کمترین تأثیر را بر روی پارامتر کیفی مدنظر داشته است.

هریک از مجموعه داده ها برای ایستگاه ورنه و گرمورد به دو بخش آموزش مدل و صحت سنجی مدل تقسیم می شوند. بنا بر مشاهدات علمی که براساس تکنیک سعی و خطا به دست آمده، بهترین تقسیم داده ها براساس درصد این گونه است که ۷۰-۸۰ درصد داده ها برای آموزش مدل و باقی داده ها برای مرحله صحت سنجی انتخاب می شوند.

در این پژوهش، داده ها برای مرحله آموزش، ۷۵ درصد داده ها را تشکیل داده و ۲۵ درصد باقی مانده برای مرحله تست یا صحت سنجی در نظر گرفته شده است. برای مدل سازی با روش های مستقل ANN، SVM و همچنین مدل ترکیبی EEMD-ANN و EEMD-SVM داده های

می یابد. آخرین IMF استخراج شده، کمترین مؤلفه فرکانسی است. با اعمال EEMD، یک سیگنال می تواند به مجموعه ای از توابع تک مؤلفه با نام توابع مُد ذاتی (IMF_s) تجزیه شود. یک تابع تک مؤلفه، تابع نوسانی نزدیک به رایج ترین و ابتدایی ترین تابع هارمونیک را نشان می دهد (علی و همکاران، ۲۰۲۰)؛ از این رو، IMF ها فرکانس هایی در محدوده بالاترین تا پایین ترین از سیگنال را شامل می شوند که به عنوان سیگنال های مدوله شده دامنه و فرکانس (AM-FM) موجود هستند، که AM پوش را حمل می کند و FM تغییر دامنه ثابت در فرکانس است و با استفاده از یک روش غربالی محاسبه می شود. برای انجام آن، یک IMF باید دو شرط را ارضا کند (هوآنگ و همکاران، ۱۹۷۱):

(۱) تعداد اکستریم (Ne) ماکزیمم و مینیمم محلی و تعداد عبور از صفرها N_z باید برابر باشند یا حداکثر یکی اختلاف داشته باشند.

$$(N_z - 1) \leq N_e \leq (N_z + 1) \quad (11)$$

(۲) در هر نقطه، مقدار میانگین پوشی که به وسیله ماکزیمم محلی و پوشی که به وسیله مینیمم محلی به وجود می آید، صفر است.

معیار ارزیابی خطا برای مقایسه عملکرد مدل ها

برای مقایسه عملکرد مدل های ارائه شده و روابط تجربی در مراحل آموزش و تست، شاخص های آماری شامل ضریب همبستگی (R)، ریشه میانگین مربعات خطا ($RMSE$)، ریشه مربعات خطا (RSE)^۲، میانگین خطای مطلق (MAE)^۳ و ریشه مطلق خطا^۴ (RAE) در معادلات (۱۲) الی (۱۶) نشان داده شده است.

$$R = \frac{\sum_{i=1}^M (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^M (O_i - \bar{O}) \sum_{i=1}^M (P_i - \bar{P})}} \quad (12)$$

$$RMSE = \frac{\sum_{i=1}^M (P_i - O_i)}{M} \quad (13)$$

$$MAE = \frac{\sum_{i=1}^M |P_i - O_i|}{M} \quad (14)$$

$$RAE = \frac{\sum_{i=1}^M |P_i - O_i|}{\sum_{i=1}^M (O_i - \bar{O})} \quad (15)$$

$$RSE = \frac{\sum_{i=1}^M (P_i - O_i)^2}{\sum_{i=1}^M (O_i - \bar{O})^2} \quad (16)$$

1- Root Mean Square Error

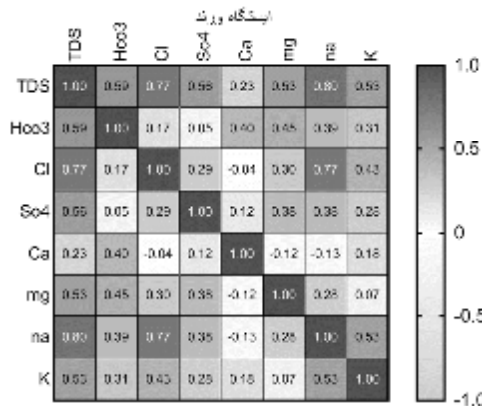
2- Root Square Error

3- Mean Absolute Error

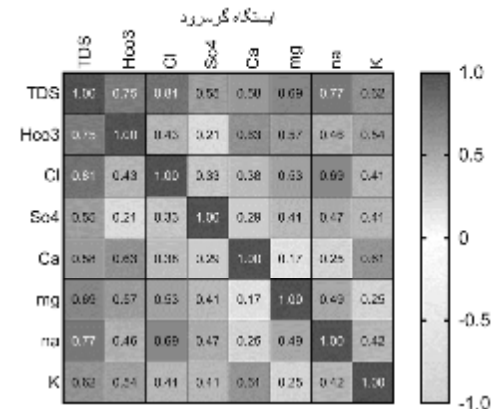
4- Root Absolute Error

5- Pearson coefficient

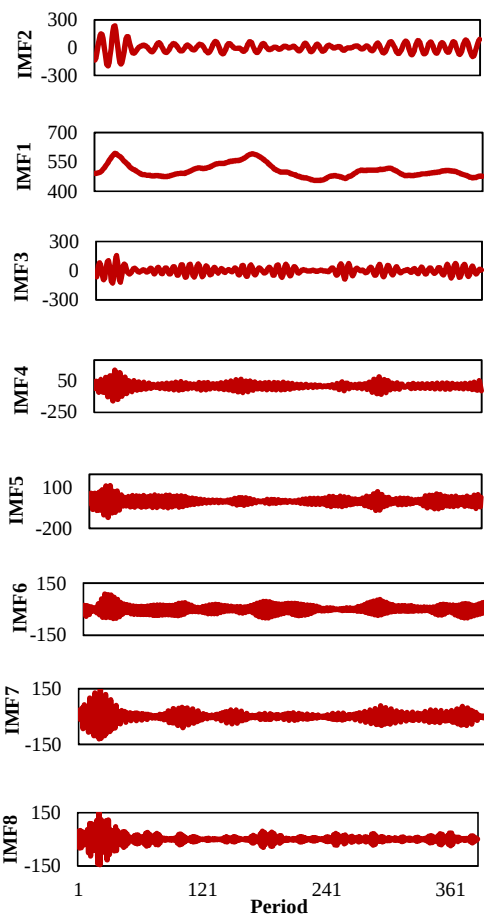
پایین (زیر سری‌های تجزیه‌شده) تبدیل شده تا نویز اضافی به‌عنوان خطا در داده‌ها و روند از داده‌های سری زمانی حذف شوند (شکل ۵). برای استخراج مقادیر TDS از ایستگاه ورنده و گرمود از الگوریتم تجزیه مد تجزیه در این تحقیق از کد موجود در برنامه Matlab R2020a استفاده شده است.



ورودی و خروجی به مدل معرفی شده و برنامه برای پیش‌بینی مقادیر TDS در ایستگاه ورنده و گرمود شروع به کار می‌کند؛ اما برای مدل‌های هوشمند مصنوعی بر پایه روش‌های پیش‌پردازش‌کننده، هریک از این پارامترها با استفاده از تکنیک پیش‌پردازش‌کننده EEMD، قبل از وارد شدن به مدل، باید به چند سیگنال فرکانس بالا و



شکل ۴- نمودار ماتریس همبستگی پیرسون برای پارامترهای فیزیکی-شیمیایی کیفیت آب برای پیش‌بینی TDS در رودخانه تجن



شکل ۵- سیگنال‌های تجزیه‌شده پارامتر TDS با استفاده از الگوریتم پیش‌پردازش‌کننده EEMD

جدول ۱- پارامترهای آماری مجموعه داده‌های کیفی آب برای ایستگاه ورنده و گرمود

ایستگاه	متغیر	محدوده	چولگی	انحراف معیار
گرمود	Hco3	0/6 - 15/5	3/96	1/06
	Cl	0/2 - 12/2	2/01	1/34
	So4	0 - 4/33	0/91	0/68
	Ca	0/6 - 8/1	1/11	0/82
	Mg	0/2 - 5/5	0/64	0/82
	Na	0/1 - 28/2	6/76	1/89
	K	0/01 - 0/4	4/91	0/02
	TDS	110 - 2024	4/02	130/57
ورنده	Hco3	1/5 - 22	5/54	1/38
	Cl	0/2 - 29/5	4/13	2/81
	So4	0/03 - 9/4	2/85	0/87
	Ca	1 - 6	0/62	0/86
	Mg	0/5 - 8/2	2/09	0/86
	Na	0/2 - 26	3/93	3/21
	K	0/01 - 0/2	1/37	0/024
	TDS	31 - 2430	3/45	254/05

نتایج مدل سازی پارامتر TDS در ایستگاه ورنه

در این بخش به تجزیه و تحلیل نتایج استخراج شده از مدل های پیشنهادی مستقل ANN، SVM و همچنین مدل ترکیبی EEMD-ANN و EEMD-SVM پرداخته می شود. جدول ۲ و ۳ نیز مقادیر شاخص های آماری خطا را برای بررسی نحوه عملکرد و دقت هریک از روش های پیشنهادی در ایستگاه ورنه ارائه می دهد. در این ایستگاه روش ANN در مرحله آموزش ضعیف ترین پیش بینی را با توجه به شاخص ریشه میانگین کمترین مربعات خطا ($RMSE=56/66$) و ریشه خطای نسبی ($RSE=1/98$) در مقایسه با دیگر روش های مستقل و ترکیبی برای پیش بینی مقادیر TDS در ایستگاه ورنه داشته است. دلیل این ضعف می تواند در این مسئله باشد که هر مدل دارای پارامترهای درونی هستند که تأثیر بسیار زیادی بر عملکرد مدل ها دارند و روش ANN نیز از این قاعده مستثنی نیست و باید پارامترهای این مدل را با سعی و خطا به دست آورد؛ بنابراین این مورد می تواند عدم قطعیت بالایی به نتایج به دست آمده از روش ANN برای پیش بینی مقادیر TDS در ایستگاه ورنه به همراه داشته باشد. در این ایستگاه، روش SVM با ترکیب روش تجزیه مد تجربی گروهی بهترین دقت ($R^2=0/98$) را در پیش بینی مقادیر TDS در رودخانه تجن داشته است.

در روش EEMD-SVM شاخص خطای مطلق میانگین برابر با ۱۴/۳۵ و شاخص خطای مطلق نسبی ۱۱/۳۲ گزارش شده که با داشتن کمترین مقادیر نشانگر قدرت

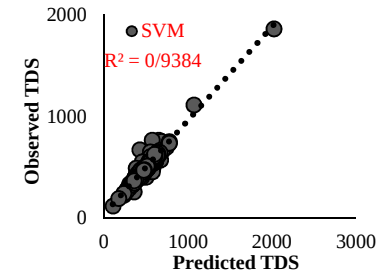
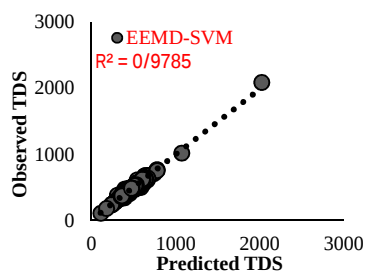
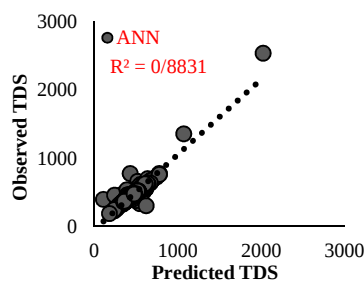
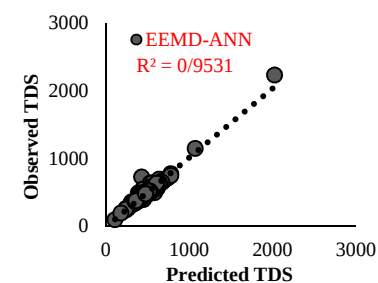
بالای آموزش در این ایستگاه بود. برای تجزیه و تحلیل بیشتر دقت و عملکرد هریک از روش های مستقل و ترکیبی در پیش بینی مقادیر TDS در مرحله آزمایش و آزمون از ایستگاه ورنه، شکل ۶ و ۷ نشان دهنده نحوه پیش بینی مدل های هوشمند مصنوعی برای دوره مورد نظر است. در آخر، درصد خطای نسبی هریک از روش های مستقل و ترکیبی را در پیش بینی مقادیر TDS در ایستگاه ورنه نسبت به مقادیر مشاهداتی در شکل ۸ می توان مشاهده کرد؛ به طوری که با توجه به نتایج، مشاهده شد که روش ترکیبی EEMD-SVM درصد خطای کمتری را نشان داد.

جدول ۲- نتایج عملکرد مرحله آموزش مدل های روش های مستقل و ترکیبی برای ایستگاه ورنه

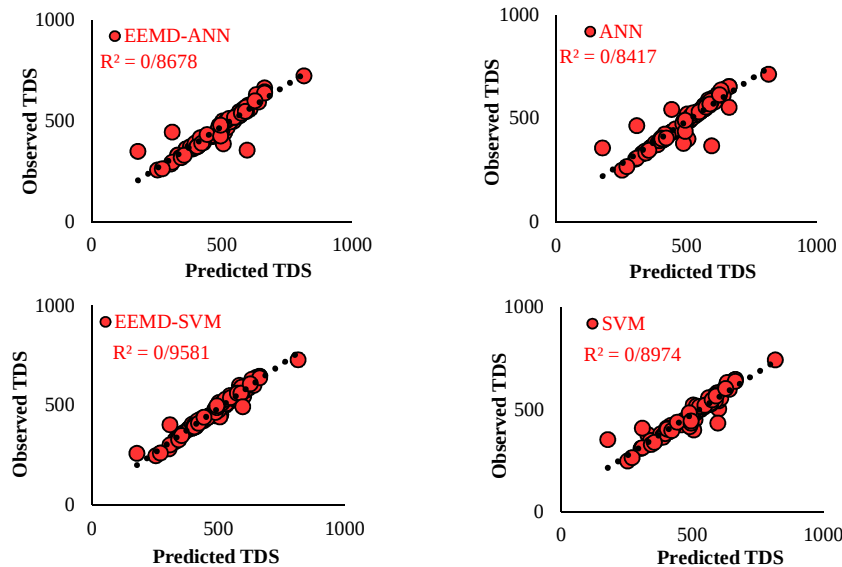
مدل	RAE	MAE	RSE	RMSE	R
ANN	۷۹/۹	۲۳/۲۱	۱/۹۸	۵۶/۶۶	۰/۹۳
EEMD-ANN	۲۴/۱۶	۱۷/۹۲	۰/۵۹	۳۱/۴۱	۰/۹۷
SVM	۲۷/۷۶	۲۳/۱۹	۰/۶۱	۳۵/۳۷	۰/۹۶
EEMD-SVM	۱۱/۳۲	۱۴/۳۵	۰/۳۱	۲۰/۲۳	۰/۹۸

جدول ۳- نتایج عملکرد مرحله آزمایش مدل های روش های مستقل و ترکیبی برای ایستگاه ورنه

مدل	RAE	MAE	RSE	RMSE	R
ANN	۲۷/۵۱	۲۲/۵۶	۰/۳۹	۴۳/۸	۰/۹۱
EEMD-ANN	۳۱/۲۸	۳۴/۴۶	۰/۴۳	۴۷/۶۸	۰/۹۳
SVM	۲۱/۶۹	۲۹/۱۳	۰/۲۷	۴۱/۰۱	۰/۹۴
EEMD-SVM	۱۱/۶۴	۱۹/۲۴	۰/۱۸	۲۷/۲۹	۰/۹۷



شکل ۶- نمودار پراکندگی مقادیر TDS مشاهداتی و محاسباتی در مرحله آموزش برای هریک از مدل های پیشنهادی در ایستگاه ورنه



شکل ۷- نمودار پراکندگی مقادیر TDS مشاهداتی و محاسباتی در مرحله آزمون برای هریک از مدل‌های پیشنهادی در ایستگاه ورنه

نتایج مدل‌سازی پارامتر TDS در ایستگاه گرمود

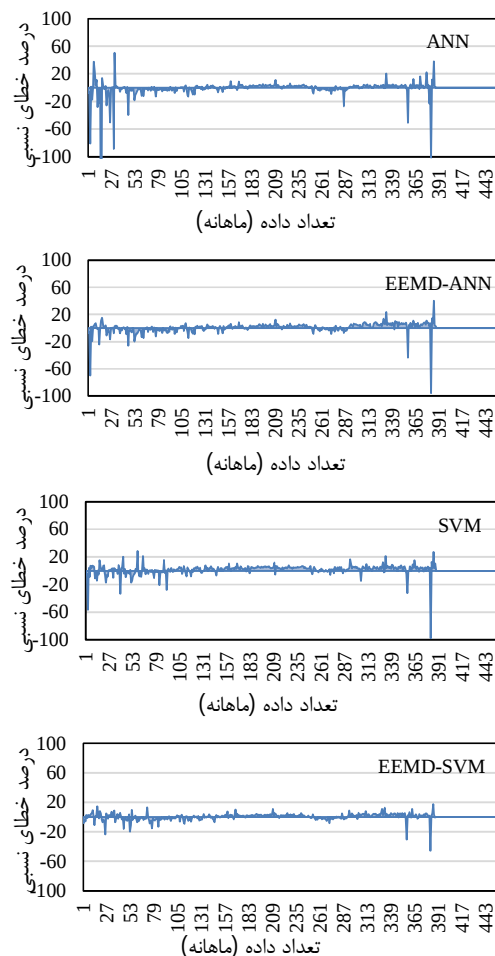
جدول‌های ۴ و ۵ نیز مقادیر شاخص‌های آماری خطا را برای بررسی نحوه عملکرد و دقت هریک از روش‌های پیشنهادی در ایستگاه گرمود را در مراحل آموزش و تست ارائه می‌دهد. در این ایستگاه روش ANN در مرحله آموزش ضعیف‌ترین پیش‌بینی را با توجه به شاخص ریشه میانگین کمترین مربعات خطا ($RMSE=87/75$) و میانگین کمترین مربعات خطا ($MAE=34/64$) در مقایسه با دیگر روش‌های مستقل و ترکیبی برای پیش‌بینی مقادیر TDS در ایستگاه گرمود داشت. در این ایستگاه، روش SVM با ترکیب روش تجزیه مد تجزیه گرووی، بهترین دقت ($R=0/98$) را در پیش‌بینی مقادیر TDS در مرحله آزمایش و دقت ($R=0/99$) را در مرحله آزمون داشت.

جدول ۴- نتایج عملکرد مرحله آموزش مدل‌های روش‌های

مستقل و ترکیبی برای ایستگاه گرمود					
مدل	RAE	MAE	RSE	RMSE	R
ANN	42/31	34/64	0/23	87/75	0/94
EEMD-ANN	19/4	37/13	0/08	66/23	0/96
SVM	22/47	54/35	0/08	77/96	0/96
EEMD-SVM	12/52	21/8	0/07	45/26	0/98

جدول ۵- نتایج عملکرد مرحله آزمایش مدل‌های روش‌های

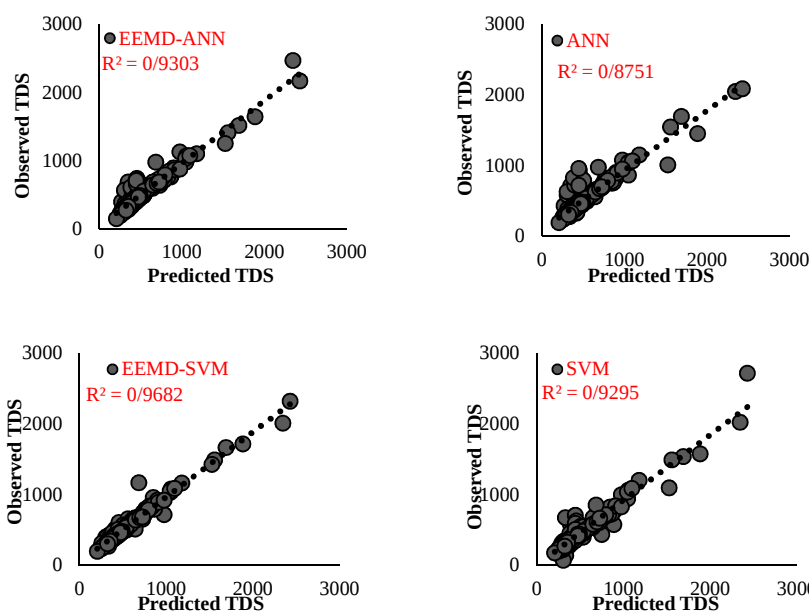
مستقل و ترکیبی برای ایستگاه گرمود					
مدل	RAE	MAE	RSE	RMSE	R
ANN	28/25	56/41	0/12	79/15	0/96
EEMD-ANN	55/13	78/29	0/36	91/35	0/96
SVM	22/1	26/32	0/17	53/59	0/97
EEMD-SVM	12/57	28/39	0/09	40/06	0/99



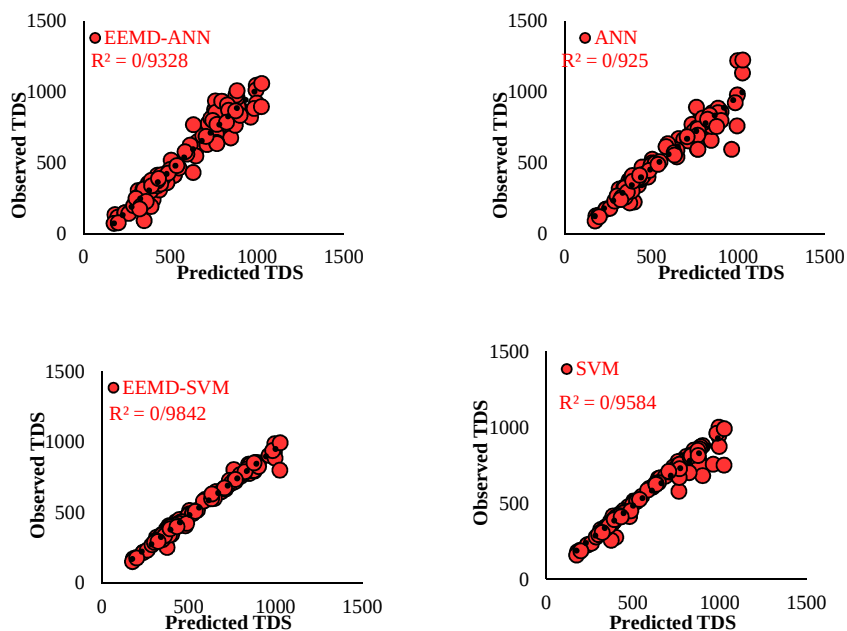
شکل ۸- نمودار سری زمانی درصد خطای نسبی هریک از مدل‌های پیشنهادی در ایستگاه ورنه

شکل ۹ و ۱۰ نشان دهنده نحوه پیش بینی مدل های هوشمند مصنوعی برای مقادیر TDS در مرحله آزمایش و آزمون از ایستگاه گرمود است. همچنین، براساس نتایج به دست آمده، این گونه به نظر می رسد که روش پیش پردازش کننده تجزیه مد تجربی در ترکیب با روش مستقل SVM کارا بوده و توانست در بیشتر پیش بینی ها

دقت قابل قبول تری ارائه دهد. در آخر، درصد خطای نسبی هریک از روش های مستقل و ترکیبی را در پیش بینی مقادیر TDS در ایستگاه گرمود نسبت به مقادیر مشاهداتی در شکل ۱۱ می توان مشاهده کرد؛ به طوری که با توجه به نتایج، مشاهده شد که روش ترکیبی EEMD-SVM درصد خطای کمتری را نشان داده است.



شکل ۹- نمودار پراکندگی مقادیر TDS مشاهداتی و محاسباتی در مرحله آموزش برای هریک از مدل های پیشنهادی در ایستگاه گرمود



شکل ۱۰- نمودار پراکندگی مقادیر TDS مشاهداتی و محاسباتی در مرحله آزمون برای هریک از مدل های پیشنهادی در ایستگاه گرمود

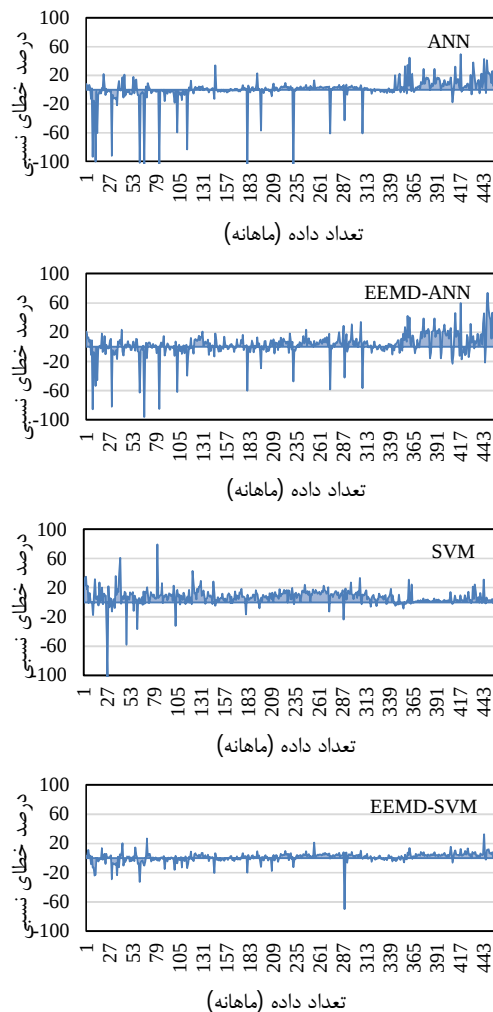
(mg, na, K)، ماه و سال برداشت درجهت پیش‌بینی مقادیر کل املاح محلول ماهانه رودخانه تجن در ایستگاه ورنه و گرمرو در نظر گرفته شد. در این سناریو، داده‌های ماهانه مقادیر کل املاح محلول ماهانه در ایستگاه ورنه و گرمرو به دو قسمت آموزش (۷۵ درصد داده‌ها) و تست (۲۵ درصد داده‌ها) تقسیم شدند.

- برای مدل‌های هوشمند مصنوعی بر پایه روش پیش‌پردازش‌کننده، هریک از پارامترهای هواشناسی با استفاده از تکنیک پیش‌پردازش‌کننده EEMD قبل از وارد شدن به مدل، باید به چند سیگنال فرکانس بالا و پایین (زیر سری‌های تجزیه‌شده) تبدیل شده تا نویز اضافی به‌عنوان خطا در داده‌ها و روند از داده‌های سری زمانی حذف شوند. نتایج نشان داد که ترکیب هریک از مدل‌های هوشمند مصنوعی با رویکرد EEMD عملکرد قابل‌قبولی از پیش‌بینی دبی ماهانه جریان رودخانه‌ها داشته است.
- نتایج مربوط به پیش‌بینی هریک از مدل‌های پیشنهادی در این تحقیق برای ایستگاه ورنه، روش مدل ANN در مرحله آموزش و آزمایش ضعیف‌ترین پیش‌بینی را با توجه به شاخص ریشه میانگین مربعات خطا ($RMSE=56/66$) و میانگین مربعات خطا ($RMSE=43/8$) در مقایسه با دیگر روش‌های مستقل و ترکیبی برای پیش‌بینی مقادیر مواد جامد محلول ماهانه رودخانه تجن در ایستگاه داشت.

همچنین نتایج مربوط به پیش‌بینی هریک از مدل‌های پیشنهادی در این تحقیق برای ایستگاه گرمرو، روش مدل ANN در مرحله آموزش ضعیف‌ترین پیش‌بینی را با توجه به شاخص ریشه میانگین مربعات خطا ($RMSE=43/8$) و ($MAE=56/41$) در مقایسه با دیگر روش‌های مستقل و ترکیبی برای پیش‌بینی مقادیر مواد جامد محلول ماهانه رودخانه تجن در ایستگاه گرمرو داشته است.

منابع

۱. امیری ا. و امیری س. ۱۳۹۴. پیش‌بینی پارامترهای کیفی رودخانه با استفاده از شبکه



شکل ۱۱- نمودار سری زمانی درصد خطای نسبی هریک از مدل‌های پیشنهادی در ایستگاه گرمرو

نتیجه‌گیری

هدف از انجام این تحقیق، استفاده از تکنیک‌های داده‌کاوی در تخمین مقادیر TDS رودخانه تجن در مقیاس ماهانه مبتنی بر الگوریتم‌های تطبیقی هوشمند با استفاده از پارامترهای فیزیکی-شیمیایی کیفی آب است. خلاصه نتایج این تحقیق در ذیل بیان شده است.

- در این تحقیق، از روش‌های هوشمند مصنوعی نظیر مدل شبکه‌های هوشمند مصنوعی و ماشین‌های بردار پشتیبان برای پیش‌بینی مقادیر کل املاح محلول ماهانه رودخانه تجن در ایستگاه ورنه و گرمرو استفاده شد. به‌طوری‌که، برای حوضه ورنه و گرمرو پارامترهای ورودی به مدل شامل آنیون‌ها (HCO_3 , Cl , SO_4)، کاتیون‌ها (Ca ,

۹. نوشادی م.، احمدزاده م.، احمدزاده م. و امین س.، ۱۳۸۶. شبیه سازی و پیش بینی کلسیم، منیزیم، سدیم، پتاسیم و سولفات در رودخانه زاینده رود با استفاده از شبکه های عصبی مصنوعی، چهارمین همایش ملی علوم و مهندسی آبخیزداری ایران مدیریت حوزه های آبخیز، کرج، ۱۲ ص.
۱۰. یوسفی م.، ۱۳۹۴. برآورد پارامترهای کیفی رودخانه های تجن و بابلرود مازندران با استفاده از مدل برنامه ریزی بیان ژن، پایان نامه کارشناسی ارشد، دانشکده کشاورزی، دانشگاه صنعتی شاهرود، ۱۸ ص.
11. Palani S. Liong S. Y. and Tkalich P. 2008. An ANN application for water quality forecasting. Marine pollution bulletin. 56(9): 1586-1597.
12. Ali M. Prasad R. Xiang Y. and Yaseen Z. M. 2020. Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. Journal of Hydrology. 584: 124647.
13. Arafa M. Alqedra M. and An-Najjar H. 2011. Neural network models for predicting shear strength of reinforced normal and high strength concrete deep beams. Journal of Applied Sciences. 11(2): 266-274.
14. Faruk D. Ö. 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering applications of artificial intelligence. 23(4): 586-594.
15. Hosseini Vaez S. R. Naderpour H. and Barati M. 1396. Prediction of flexural strength of NSM-FRP reinforced beams using artificial neural networks. Journal of Structural and Construction Engineering. 4(4): 16-28.
16. Huang N. E. Shen Z. Long S. R. Wu M. C. Shih H. H. Zheng Q. and Liu H. H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences. 454: 903-995.
17. Kaya A. 2010. Artificial neural network study of observed pattern of scour depth around bridge piers. Computers and Geotechnics. 37(3): 413-418.
18. Maier H. R. and Dandy G. C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water resources research. 32(4): 1013-1022.
19. Nehdi M. El Chabib H. and El Naggar M. H. 2001. Predicting performance of self-curing concrete. Journal of Materials in Civil Engineering. 13(4): 267-275.
۲. پارساخو آ.، ۱۳۹۱. بررسی مقدار رواناب و هدررفت خاک بخش های مختلف ساختمان جاده جنگلی با استفاده از شبیه ساز باران (مطالعه موردی: سری های لتتالار و لولت در حوزه آبخیز تجن)، پایان نامه دکتری، رشته مهندسی منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ۱۶ ص.
۳. حسنی ن. کریملو ک. و رشیدی مهرآبادی ع.، ۱۳۹۵. مدیریت بحران آب در کشورهای درحال توسعه، هشتمین کنفرانس بین المللی مدیریت جامع بحران، تهران، ۱۵ ص.
۴. رجایی ط. و رحیمی بنماران ر.، ۱۳۹۵. پیش بینی پارامترهای کیفی CL، NO3 رودخانه کرج با استفاده از مدل های ترکیبی شبکه عصبی-موجکی ANN و MLR. فصلنامه آب و خاک. ۳۰(۱): ۱۵-۲۹.
۵. عظیمی ق.، ۱۳۸۷. شناخت و ارزیابی کیفیت آب رودخانه جاجرود، پایان نامه کارشناسی ارشد مهندسی محیط زیست، دانشکده عمران، دانشگاه صنعتی شریف، ۱۲۶ ص.
۶. فراهی ف.، مسعودیان م. و درزی نفت چالی ع.، ۱۳۹۷. بررسی آلاینده های فیزیکی و شیمیایی بر کیفیت آب رودخانه تجن در محدوده شهری براساس رهنمود WFD، پنجمین همایش ملی تغییر اقلیم و گاهشناسی درختی در اکوسیستم های خزری، ساری، ۱۸ ص.
۷. کرمی م.، کاشفی پور م.، معاضد ه. و فروغی ح.، ۱۳۸۵. پیش بینی کیفیت آب رودخانه کارون با استفاده از شبکه های عصبی مصنوعی، هفتمین سمینار بین المللی مهندسی رودخانه، اهواز، ۱۶ ص.
۸. میثاقی ف. و محمدی ک.، ۱۳۸۳. پیش بینی تغییرات کیفیت آب رودخانه زاینده رود با استفاده از شبکه های عصبی مصنوعی، دومین کنفرانس ملی دانشجویی منابع آب و خاک، شیراز، ۱۴ ص.

- assisted data analysis method. *Advances in adaptive data analysis*. 1(01): 1-41.
30. Xiang Z. Yan J. and Demir I. 2020. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water resources research*. 56(1): e2019WR025326.
 - compacting concrete mixtures using artificial neural networks. *Materials Journal*. 98(5): 394-401.
 20. Pham A. D. Hoang N. D. and Nguyen Q. T. 2016. Predicting compressive strength of high-performance concrete using metaheuristic-optimized least squares support vector regression. *Journal of Computing in Civil Engineering*. 30(3): 06015002.
 21. Rezaie-Balf M. and Kisi O. 2018. New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. *Hydrology Research*. 49(3): 939-953.
 22. Rezaie-Balf M. Attar N. F. Mohammadzadeh A. Murti M. A. Ahmed A. N. Fai C. M. and El-Shafie A. 2020. Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: Comparative assessment of a noise suppression hybridization approach. *Journal of Cleaner Production*. 271: 122576.
 23. Rezaie-Balf M. Fani Nowbandegani S. Samadi S. Z. Fallah H. and Alaghmand S. 2019. An Ensemble Decomposition-Based Artificial Intelligence Approach for Daily Streamflow Prediction. *Water*. 11(4): 709-729.
 24. Schölkopf B. and Smola A. J. 2002. *Learning with kernels: support vector machines, Regularization, Optimization, and beyond*. MIT press. 1(2): 321-358.
 25. Sharafati A. Asadollah S. B. H. S. and Hosseinzadeh M. 2020. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Safety and Environmental Protection*. 140: 68-78.
 26. Singh K. P. Basant A. Malik A. and Jain G. 2009. Artificial neural network modeling of the river water quality—a case study. *Ecological Modelling*. 220(6): 888-895.
 27. Tang J. Gao F. Liu F. and Chen X. 2020. A denoising scheme-based traffic flow prediction model: Combination of ensemble empirical mode decomposition and fuzzy C-Means neural network. *IEEE Access*. 8: 11546-11559.
 28. Tomić A. Š. Antanasijević D. Ristić M. Perić-Grujić A. and Pocajt V. 2018. Application of experimental design for the optimization of artificial neural network-based water quality model: a case study of dissolved oxygen prediction. *Environmental Science and Pollution Research*. 25(10): 9360-9370.
 29. Wu Z. and Huang N. E. 2009. Ensemble empirical mode decomposition: a noise-

