

ارزیابی تأثیر روش های پیش پردازش کننده در عملکرد تکنیک های محاسبات نرم در تخمین دمای نقطه شبنم

محمد تیموری^۱، علیرضا قائمی^۲، غلامرضا عزیزیان^۳، سید آرمان هاشمی منفرد^{۴*}

^۱ دانشجوی کارشناسی ارشد گروه مهندسی عمران گرایش مهندسی و مدیریت منابع آب، دانشگاه سیستان و بلوچستان

^۲ دانشجوی دکتری گروه مهندسی عمران گرایش مهندسی و مدیریت منابع آب، دانشگاه سیستان و بلوچستان

^۳ دانشیار گروه مهندسی عمران، دانشگاه سیستان و بلوچستان، زاهدان

^۴ دانشیار گروه مهندسی عمران، دانشگاه سیستان و بلوچستان، زاهدان

تاریخ دریافت: ۹۸/۱۲/۱۸، تاریخ پذیرش: ۹۹/۰۳/۲۳

چکیده

نقطه شبنم دمایی است که در آن، هوا با استفاده از بخار آب، اشباع شده است. در واقع در اثر کاهش دما در فشار ثابت، به دمایی می‌رسیم که در آن، هوا توسط بخار آب اشباع شده است، به این دما، دمای نقطه شبنم می‌گویند. از طرفی، با توجه به کمبود آب در جهان و تعداد محدود ایستگاه های اندازه گیری نقطه شبنم، استفاده از روش های مختلف برای تخمین دمای نقطه شبنم با دقت قابل قبول مورد نیاز است. از طرفی، با توجه به تعداد محدود ایستگاه های اندازه گیری نقطه شبنم، استفاده از روش های مختلف برای تخمین دمای نقطه شبنم با دقت قابل قبول مورد نیاز است. امروزه بکارگیری روش های هوشمند با توجه به توانایی آن‌ها در برقراری ارتباط بین ورودی و خروجی‌ها، مورد توجه بسیاری از محققین قرار گرفته‌اند. در این تحقیق، عملکرد دو روش ماشین بردار پشتیبان (SVM) و برنامه ریزی بیان ژن (GEP) برای پیش بینی دمای نقطه شبنم بررسی شد. همچنین از تبدیل موجک (WT) به عنوان یک روش پیش پردازش کننده داده‌ها به منظور بهبود نتایج استفاده شد. ارزیابی عملکرد مدل‌ها با شاخص های آماری خطا نشان داد که مدل‌های ترکیبی عملکرد بهتری نسبت به مدل‌های منفرد داشتند. همچنین، مدل ترکیبی حاصل از موجک با روش برنامه نویسی بیان ژن (GEP^W) در مقایسه با ماشین بردار پشتیبان (SVM^W) از دقت کمتری برخوردار بود و در نهایت در میان مدل‌های در نظر گرفته شده، مدل SVM^{W-Haar} با بالاترین دقت (NSE=۰/۸۹ و RMSE=۰/۸۵) توانست مقادیر دمای نقطه شبنم را تخمین بزند و به عنوان مدل برتر انتخاب گردید.

کلمات کلیدی: دمای نقطه شبنم، ماشین بردار پشتیبان، برنامه نویسی بیان ژن، تبدیل موجک

امروزه کمبود آب به یک چالش جهانی تبدیل شده است که کشورهای مختلف را درگیر خود کرده است. با توجه به افزایش دمای کره زمین و مصرف بی رویه آب، منابع آب شیرین دنیا رو به کاهش است و در کشورهایی همچون ایران که دارای اقلیم خشک و نیمه خشک هستند به شدت نمود پیدا کرده است. شبنم یکی از حالت های آب در چرخه هیدرولوژی است که می توان از آن استفاده های مفیدی در جهت حل معضل کم آبی داشت. دمای نقطه شبنم، دمایی است که رطوبت موجود در هوا در اثر سرد شدن دمای محیط تغییر فاز داده و به شکل قطرات آب در می آید و در صورتی که این دما بیشتر کاهش یابد، این قطرات بروی سطح اجسام شکل می گیرند. دانستن دقیق دمای نقطه شبنم، در علوم همچون هیدرولوژی، هواشناسی، کشاورزی بیشتر مورد توجه قرار گرفته است زیرا می توان از فواید آن در جهت تامین آب استفاده و از آسیب های شبنم مانند یخ زدگی و موج های گرم پیشگیری کرد. در دمای نقطه شبنم مقدار فشار اشباع با مقدار فشار واقعی برابر است. این دما بیانگر میزان رطوبت موجود در هوا می باشد و با استفاده از آن می توان میزان رطوبت نزدیک به سطح زمین را تخمین زد. این موضوع در مناطقی با الگوهای بارش نامنظم مانند مناطق خشک و نیمه خشک که دارای بارش کم هستند، از درجه اهمیت بیشتری برخوردار است زیرا در این مناطق به دلیل کمبود منابع آب زیر زمینی، آب های سطح زمین و رطوبت موجود در هوا می تواند نیاز گیاهان و حیوانات را تامین کند و شبنم می تواند بخشی از این رطوبت باشد. از طرفی با اقدامات کنترلی مناسب می توان از یخ زدگی محصولات جلوگیری کرد. پس می توان نتیجه گرفت که دمای نقطه شبنم به صورت مستقیم در زندگی انسان ها، حیوانات و تولیدات محصولات کشاورزی تاثیر گذار است. از این رو، با تخمین درستی از این پارامتر می توان به مدیریت و اقدامات کنترلی مناسبی در زمینه های مختلفی مانند هیدرولوژی، هواشناسی، کشاورزی دست یافت. (محمدی و همکاران، ۲۰۱۵، مهدی زاده و همکاران، ۲۰۱۷، امیر

مجاهدی و همکاران، ۲۰۱۶، ذونعمت کرمانی، ۲۰۱۲، دانگ و همکاران، ۲۰۲۰، ناگانا و همکاران، ۲۰۱۹، داکو همکاران، ۲۰۱۸).

در دهه های اخیر، روش های محاسبات نرم (Soft computing methods)، کاربردهای فراوانی پیدا کرده است و جایگزین روش های سنتی و مدل های کلاسیک شده اند. در واقع این روش ها، تکنیک های محاسباتی هستند که عمدتاً برای طبقه بندی، پیش بینی و تشخیص الگو ها بکار می روند. از طرفی، با توجه به پیشرفت علم در زمینه های مختلف و کاربرد فراوان انواع این روش ها مانند شبکه های عصبی مصنوعی (Artificial neural networks)، سیستم تطبیقی استنتاج عصبی فازی (Adaptive neuro-fuzzy inference system)، برنامه بیان ژن (Gene-expression programming) و ماشین بردار پشتیبان (Support vector machine) کمک شایانی در جهت حل مسائل گوناگون در زمینه مهندسی محیط زیست کرده اند (کیسی و همکاران، ۲۰۱۳، قائمی و همکاران، ۲۰۱۹، منتظری و همکاران، ۲۰۱۸، کیم و همکاران، ۲۰۱۹، دئو و همکاران، ۲۰۱۶، علیزاده و همکاران، ۲۰۱۸، شماری و همکاران، ۲۰۱۶). با توجه به اهمیت دمای نقطه شبنم، روش های مختلفی از محاسبات نرم که در آن ها از پارامترهای هیدرولوژی مختلفی به عنوان ورودی در نظر گرفته شده اند، استفاده شده است (باغبان و همکاران، ۲۰۱۶، شیری، ۲۰۱۸، فتح الله زاده عطار و همکاران، ۲۰۱۸، قاسم و همکاران، ۲۰۱۹، دانگ و همکاران، ۲۰۲۰). بطور مثال، محمدی و همکاران (۲۰۱۵) عملکرد روش های ماشین یادگیری افراطی (Extreme learning machine)، ANN و SVM به منظور تخمین دمای نقطه شبنم را ارزیابی کردند. نتایج آنها حاکی از برتری مدل ELM نسبت به دو روش دیگر بود. روش سیستم تطبیقی استنتاج عصبی فازی نیز توسط محمدی و همکاران (۲۰۱۶) جهت تخمین دمای نقطه شبنم انتخاب گردید. معیارهای ارزیابی دلالت بر عملکرد قابل قبول مدل در نظر گرفته شده در تخمین این پارامتر مهم داشتند. قاسم و همکاران (۲۰۱۹) برای تخمین

شوند. همچنین، مقایسه روش های مختلف می تواند راهی برای تعیین بهترین مدل برای یک سناریوی مشخص باشد (ابهیناو و همکاران، ۲۰۱۷، خسروی و همکاران، ۲۰۱۸، جراف و همکاران، ۲۰۱۹، یو و همکاران، ۲۰۲۰، قائمی و همکاران، ۲۰۲۱) با توجه به موارد ذکر شده، هدف تحقیق حاضر توسعه یک مدل یکپارچه دقیق برای تخمین دمای نقطه شبنم در زاهدان است. بدین منظور، عملکرد مدل های حاصل از ترکیب یکی از روش های پیش پردازش کننده بنام موجک و مدل های پیش بینی کننده ماشین بردار پشتیبان و برنامه نویسی بیان ژن برای تخمین دمای نقطه شبنم ارزیابی شده است.

مواد و روش ها

منطقه مورد مطالعه

استان سیستان و بلوچستان در جنوب شرقی ایران واقع شده است و جزء مناطق خشک ایران به شمار می آید (شکل ۱). شهر زاهدان به عنوان مرکز استان در شمال این استان واقع شده است و منطقه ای با بارش کم و نامنظم می باشد و آب مورد نیاز این شهر از شهر شمالی استان بنام زابل و سفره های زیر زمینی تامین می گردد. منابع آب شمال استان از افغانستان آگیری می شود که منجر به وابستگی بخش کشاورزی استان به این کشور شده است (لشکری پور، ۲۰۰۵).

دمای نقطه شبنم از مدل درخت (Model tree) بهره بردند. آن ها از داده های فشار واقعی بخار، دمای میانگین، رطوبت نسبی، سرعت باد و تابش روزانه خورشید در یک دوره ۱۸ ساله بعنوان ورودی در نظر گرفتند. نتایج نشان داد که مدل درخت عملکرد خوبی را از خود نشان داده است. فتح الله زاده عطار و همکاران (۲۰۱۸) برای تخمین دمای نقطه شبنم از سه روش اسپلاین رگرسیون تطبیقی چندگانه (Multivariate adaptive regression splines)، GEP و SVM استفاده کرده اند. آن ها ۱۳ ایستگاه در مناطق خشک ایران را بررسی کردند بطوریکه ۹ ایستگاه را برای آموزش و ۴ ایستگاه را برای تست در نظر گرفتند. طبق نتایج حاصل از شاخص های آماری خطا، مدل MARS عملکرد بهتری نسبت به مدل های دیگر داشت. از طرفی، روش های حاضر جهت تخمین دمای نقطه شبنم ممکن است دارای مزایا و معایبی باشند. از این رو، روش های جدید باید این معایب را برطرف کنند. با توجه به محدودیت هر کدام از مدل ها، توانایی روش های پیش بینی کننده بطور قابل ملاحظه ای بستگی به ساختار مدل و متغیرهای ورودی دارد. از این رو، ترکیب برخی از روش ها می تواند به بهبود دقت منجر شود. روش های ترکیبی به طور کلی به عنوان یک تکنیک یکپارچه ارائه می شوند که در آنها الگوریتم های تقویت کننده و پیش پردازش را جهت افزایش دقت مدل ها بکار گرفته می



شکل ۱: موقعیت منطقه مورد مطالعه

جدول ۱. ویژگی های آماری پارامترهای مورد استفاده در دوره اقلیمی انتخابی (۲۰۰۰ - ۲۰۱۸)

| پارامتر | واحد | میانگین | بیشترین مقدار | کمترین مقدار | انحراف معیار |
|-----------|--------------------|---------|---------------|--------------|--------------|
| T_{max} | درجه سلسیوس (°C) | ۲۷/۶۵ | ۴۳/۴ | -۳/۸ | ۸/۵۸ |
| T_{min} | درجه سلسیوس (°C) | ۱۱/۲۶ | ۲۹/۴ | -۱۳ | ۸/۳۱ |
| S_h | ساعت (hour) | ۹/۳۳ | ۱۳/۱ | ۰ | ۲/۶۱ |
| W_S | متر بر ثانیه (m/s) | ۳/۳۳ | ۱۵/۲۵ | ۰ | ۱/۶۳ |
| R_H | درصد (%) | ۲۶/۷۸ | ۹۹ | ۴/۱۲ | ۱۶/۳۵ |
| TD | درجه سلسیوس (°C) | -۲/۹۱ | ۱۷/۴۶ | -۳۸/۱۲ | ۵/۸۹ |

انتخاب داده ها

ارائه می دهد. تحلیل موجکی قادر به نمایش جنبه های مختلف داده های متفاوت، نقاط شکست و ناپیوستگی ها می باشد که ممکن است دیگر روش های آنالیز سیگنال آنها را نشان ندهند. یکی از مزایای استفاده از این روش، تجزیه سیگنال سری زمانی به زیر سیگنال های مجزاست. WT به دو نوع تبدیل موجک پیوسته (CWT) و تبدیل موجک گسسته (DWT) تقسیم می شود که به طور کلی، DWT برای حل مسائل عملی در زمینه هیدرولوژی مفیدتر از CWT است. علاوه بر این، DWT می تواند مجموعه داده های اصلی سری زمانی را به تقریب (ai با مقیاس بالا با فرکانس پایین) و جزئیات (قطعات با فرکانس بالا و کم مقیاس) در سطوح مختلف تجزیه کند (رضایی و همکاران، ۲۰۱۷، دئو و همکاران، ۲۰۱۷).

تبدیل موجک

برنامه ریزی بیان ژن
کوزا در سال ۱۹۹۲ با توسعه دادن الگوریتم وراثتی (Genetic algorithm)، برنامه نویسی وراثتی (Genetic programming) که به منظور حل مسائل مختلف از اصول انتخاب طبیعی داروین استفاده می کند، را ارائه داد. تفاوت اصلی بین برنامه نویسی وراثتی و الگوریتم وراثتی در نشان دادن نتایج است. بطوریکه، در الگوریتم GP افراد دارای رشته های غیرخطی در اندازه و شکل می باشند در حالیکه در الگوریتم GA، افراد دارای رشته های خطی با طول ثابت می باشند. همچنین، ساختار و ساختمان

بر اساس مطالعات قبلی مانند (شیری، ۲۰۱۸، محمدی و همکاران، ۲۰۱۵، باغبان و همکاران، ۲۰۱۶). در این پژوهش، از متغیرهای دمای حداکثر (T_{max})، دمای حداقل (T_{min})، ساعات تابش خورشید (S_h)، سرعت باد (W_S)، رطوبت نسبی (R_H) و دمای نقطه شبنم (TD) که به صورت روزانه و طی یک دوره ۱۹ ساله (۲۰۰۰ - ۲۰۱۸) از ایستگاه هواشناسی زاهدان جمع آوری شده اند، استفاده شده است. از میان داده های موجود، حدود ۷۵ درصد داده ها به عنوان داده های آموزش و ۲۵ درصد باقیمانده به عنوان داده های تست در نظر گرفته شده اند. ویژگی های آماری داده های استفاده شده در جدول ۱ قابل مشاهده می باشد.

تئوری موجکی یکی از روش های علم ریاضی است که ایده اصلی آن برگرفته از تبدیل فوریه است که در قرن نوزدهم مطرح شده است. تبدیل موجک یکی از تبدیل های ریاضی کارآمد در زمینه پردازش سیگنال است. موجک ها توابع ریاضی هستند که شکل مقیاس-زمان از سری های زمانی و روابط آنها جهت آنالیز سری های زمانی که شامل متغیرها و غیرثابت ها می باشد را ارائه می دهد. تحلیل موجکی استفاده از فاصله های زمانی طولانی مدت را برای اطلاعات دارای فرکانس پائین و تناوب های کوتاهتر را برای اطلاعات دارای فرکانس بالا

بهبود می یابد و در صورت دستیابی به نتیجه مطلوب و یا رسیدن به شرایط توقف، تکامل پایان یافته و بهترین نتیجه ارائه می گردد. (زمانی و همکاران، ۲۰۱۵، کاظمی و همکاران، ۲۰۲۱).

ماشین بردار پشتیبان

ماشین بردار پشتیبان از جمله روش های یادگیری با ناظر است که برای طبقه بندی و رگرسیون بکار برده می شود و در سال ۱۹۹۲ توسط وپنیک ریاضیدان روسی و چروننکیس بر پایه تئوری یادگیری آماری ارائه شد. این روش از یک ابر صفحه خطی تفکیک گر بهینه برای دسته بندی داده ها استفاده می کند که این ابر صفحه از حداکثر کردن حاشیه استفاده می کند. به همین دلیل باید یک مساله بهینه سازی غیر خطی مقید حل شود. SVM از جمله گروه دسته بندی کننده هایی است که بدون دانستن مدل توزیع داده ها، دسته بندی را انجام می دهد و نیازی به دانستن ساختار کلی از مساله ندارد. این روش بردارهای ورودی را با استفاده از یک تابع کرنل به فضای بزرگتری برده تا بتواند تفکیک پذیری خطی را بهتر انجام بدهد. در یک مدل رگرسیونی SVM لازم است که وابستگی تابعی متغیر وابسته y به مجموعه ای از متغیرهای مستقل x تخمین زده شود. مانند دیگر مسایل رگرسیونی، فرض بر این است که رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین $f(x)$ بعلاوه یک مقدار اضافی نویز مشخص شود (لی و همکاران، ۲۰۲۰، فنچ و همکاران، ۲۰۲۰):

$$y = f(x) + \text{noise} \quad (1)$$

بنابراین پیدا کردن فرم تابع $f(x)$ از اهمیت زیادی برخوردار است که بتواند بصورت صحیح، مواردی که SVM تجربه نکرده است را پیش بینی کند. این تابع بوسیله آموزش مدل SVM بر روی مجموعه داده های آموزش که شامل فرایند بهینه سازی دائمی تابع خطا است، قابل دسترسی است. براساس تعریف این تابع خطا، دو نمونه از

الگوریتم GP به صورت سیستم درختی مدل می شود، در حالی که الگوریتم GA به صورت سیستم دو دویی (باینری) و رشته ای از اعداد حقیقی نشان داده می شود. از این رو، فریرا در سال ۱۹۹۹ برنامه نویسی بیان ژن که از توسعه GA و GP پدید آمده است را ارائه داد. در GEP، افراد به عنوان رشته های خطی با طول ثابت (کروموزوم) کدگذاری می شوند و پس از آن به عنوان رشته های غیر خطی در اندازه و شکل هایشان بیان می گردند. از مزایای روش GEP نسبت به دیگر روش ها می توان به طبیعت منحصر به فرد چند ژن بودن اشاره کرد که زمینه ارزیابی مدل های پیچیده ای را که شامل چندین زیر مدل می باشد، فراهم می آورد. همچنین، در این مدل ابتدا ساختار بلوک ها مانند مجموعه توابع، متغیرهای ورودی و هدف تعریف و سپس ساختار بهینه مدل و ضرایب در طول فرایند مدلسازی تعیین می شوند. درحالی که در دیگر روش ها مانند شبکه عصبی، در مرحله اول ساختار مدل تعیین می شود و ضرایب مدل در طول فرایند مدلسازی محاسبه می گردند (کابلی و همکاران، ۲۰۱۷، مهدی زاده و همکاران، ۲۰۲۰).

بطور کلی، در GEP، ابتدا یک جمعیت اولیه تصادفی از فرمول ها ایجاد می شود که این فرمول ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول ها) و ترمینال ها (متغیرهای مسئله و اعداد ثابت) ایجاد می شوند. در مرحله بعد، هر یک از افراد جمعیت مذکور با استفاده از توابع برازش و چرخ رولت مورد ارزیابی قرار می گیرند. منطق مکانیزم چرخ رولت به این صورت است که هر یک از کروموزوم ها بسته به میزان مناسب بودنشان (براساس تابع برازش) احتمال انتخاب شدن را دارند. به عبارت دیگر، هر چه یک کروموزوم تأثیر بهتری داشته باشد، احتمال بیشتری وجود دارد که برای تولید در نسل بعدی انتخاب شود. سپس، افراد انتخاب شده با استفاده از یکی از عملگرهای جهش و همبری، جمعیت جدیدی ایجاد می کنند و دوباره مورد ارزیابی قرار می گیرند. این فرآیند برای چندین نسل تکرار می شود و با پیشرفت نسل به جلو کیفیت جمعیت نیز به طور نسبی

مدل های SVM شناخته شده است که عبارتند از:

الف- مدل های رگرسیونی SVM نوع اول که به مدل های v-SVM مشهورند.

ب - مدل های رگرسیونی SVM نوع دوم که به مدل های ε-SVM معروف هستند.

در این تحقیق، از مدل های رگرسیونی نوع دوم که کاربرد گسترده تری دارند، استفاده شده است. برای این مدل، تابع خطا بصورت زیر تعریف می شود:

$$(2) \quad \frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i'$$

تابع خطای زیر باید با توجه به محدودیت های زیر کمینه بشود:

$$(3) \quad \begin{aligned} W^T \phi(X_i) + b - y_i &\leq \varepsilon + \xi_i' \\ y_i - W^T \phi(X_i) - b &\leq \varepsilon + \xi_i \\ \xi_i &\& \xi_i' \geq 0, i=1,2,\dots,N \end{aligned}$$

که C ثابت گنجایش، W بردار ضرایب، W^T ترانزاده بردار ضرایب، ξ_i' و ξ_i ضرایب کمبود، b ضریب ثابت، N الگوی آموزش مدل و ϕ تابع کرنل است.

توابع کرنل قابل استفاده در این مدل به صورت های گوناگون وجود دارند که دو مورد از آنها بصورت زیر بیان می شوند (کیسی و همکاران، ۲۰۱۵، نوری و همکاران، ۲۰۰۸):

الف) کرنل چندجمله ای (Polynomial) به فرم:

$$(4) \quad k(x_i, x_j) = (x_i^T x_j + \gamma)^d$$

ب) کرنل تابع پایه شعاعی (Radial basis function) به فرم:

$$(5) \quad k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

مقایسه عملکرد مدل ها در پیش بینی

در این قسمت، به منظور مقایسه عملکرد مدل های ارائه شده در مراحل آموزش و تست، از تعدادی شاخص آماری شامل ضریب همبستگی (Correlation coefficient)، ریشه میانگین مربعات خطا (Root mean square error)، ضریب خطای نش (Nash Sutcliffe coefficient) و شاخص کلی عملکرد مدل (Overall index) که بصورت زیر بیان می گردند، استفاده شده است (گندمی و همکاران، ۲۰۱۰).

$$(6) \quad R = \frac{\sum_{i=1}^N (TD_p^i - \overline{TD_p})(TD_o^i - \overline{TD_o})}{\sqrt{\left(\sum_{i=1}^N (TD_o^i - \overline{TD_o})^2 \sum_{i=1}^N (TD_p^i - \overline{TD_p})^2\right)}}$$

$$(7) \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (TD_o^i - TD_p^i)^2}{N}}$$

$$(8) \quad NSE = 1 - \frac{\sum_{i=1}^N (TD_o^i - TD_p^i)^2}{\sum_{i=1}^N (TD_o^i - \overline{TD_o})^2}$$

$$(9) \quad OI = \frac{1}{2} \left(2 - \frac{RMSE}{TD_o^{MAX} - TD_o^{MIN}} - \frac{\sum_{i=1}^N (TD_o^i - TD_p^i)^2}{\sum_{i=1}^N (TD_o^i - \overline{TD_o})^2} \right)$$

که در این روابط TD_p^i برابر با مقدار دمای نقطه شبنم پیش بینی شده، $\overline{TD_p}$ برابر با میانگین دمای نقاط شبنم پیش بینی شده، TD_o^i برابر با مقدار دمای نقطه شبنم مشاهده ای، $\overline{TD_o}$ برابر با میانگین دمای نقاط شبنم مشاهده ای، N تعداد داده ها و TD_o^{MAX} و TD_o^{MIN} به ترتیب برابر با مقدار دمای نقطه شبنم مشاهده ای کمینه و بیشینه می باشد. ضریب همبستگی در بازه ۱ و -۱ تغییر میکند و بدین صورت است که هر چه این ضریب به یک نزدیکتر باشد دقت مدل بالاتر است. شاخص RMSE از جنس پارامترهای خروجی می باشد و مقادیر بین صفر تا عدد مثبت را در بر می گیرد که هر چه به صفر نزدیکتر

شود. پس از انجام سعی و خطاهای متعدد، بهترین مقادیر انتخاب شد که در جدول ۲ آورده شده است.

ارزیابی روش برنامه نویسی بیان ژن

نتایج ارزیابی روش های GEP و GEP^W با استفاده از شاخص های اماری در دو مرحله آموزش و تست به ترتیب در جدول ۳ نشان داده شده است. مطابق با جدول ۲، در مرحله آموزش، روش GEP^{W-Coif} با کمترین مقدار خطا (RMSE=۲/۳۳) و بیشترین دقت (NSE=۰/۷۹۵ و R=۰/۹۲) بهترین عملکرد را از خود نشان داد. در مرحله تست، روش GEP^{W-Coif} با ۰/۸۵۳ NSE=۰ و ۰/۸۹۹ OI=۰ توانست مقادیر TD را بهتر از دیگر روش ها مانند GEP (OI=۰/۸۳۹ و NSE=۰/۷۵۳) پیش تخمین بزند. همچنین، مقادیر شاخص های خطا برای روش GEP و GEP^W در مرحله آموزش و تست

باشد یعنی مدل دقت بالاتری دارد. در انتها دو شاخص NSE, OI اعداد کوچکتر از یک را اختیار می کند که هر چه به ۱ نزدیکتر باشند یعنی عملکرد مدل بهتر است.

نتایج و بحث

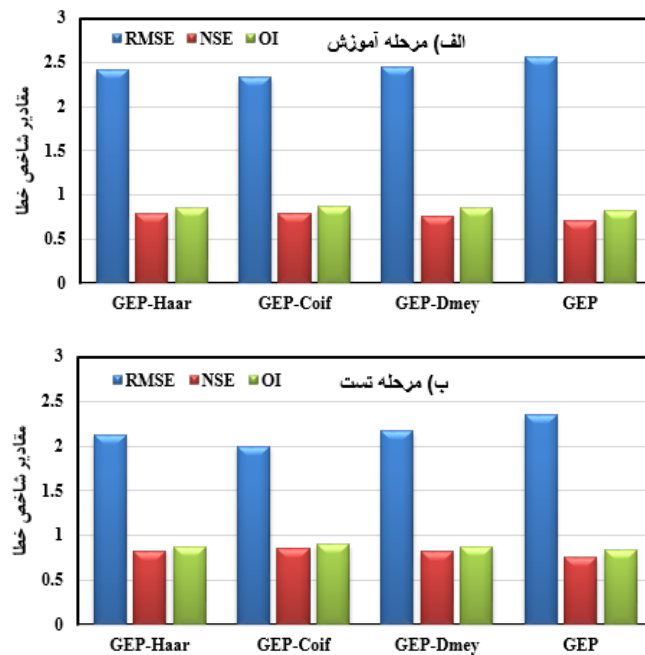
در این قسمت با استفاده از شاخص های اماری عملکرد هر یک از روش ها به صورت مجزا در دو مرحله آموزش و تست ارزیابی خواهد شد. علاوه بر این، بمنظور بررسی عملکرد روش پیش پردازش WT از موجک های مادر Haar, Coif, Dmey استفاده شده است. ماشین بردار پشتیبان از کرنل های مختلفی برای اهداف خود استفاده می کند. دو کرنل چندجمله ایی (Poly) و تابع پایه شعاعی (RBF) برای پیش بینی دمای نقطه شبنم بکار گرفته شد. در مدل GEP، پارامترهای اولیه و معیار توقف باید مشخص شوند تا با استفاده از آنها عملیات متوقف

جدول ۲- پارامترهای مربوط به برنامه نویسی بیان ژن

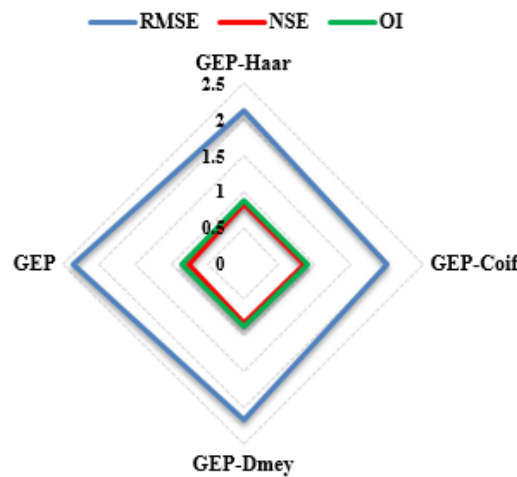
| تنظیمات | تشریح پارامتر | ردیف |
|---|-------------------|------|
| $+, -, \times, \div, x^2, x^3, \frac{1}{x}, \sqrt{x}, \sqrt[3]{x}, \log x, \ln x$ | مجموعه توابع | ۱ |
| ۳ | تعداد ژن ها | ۲ |
| ۳۰ | تعداد کروموزوم | ۳ |
| ۵ | حداکثر طول عبارات | ۴ |

جدول ۳. ارزیابی عملکرد مدل های GEP در مرحله آموزش و تست

| OI | NSE | RMSE | R | Model |
|-------------|-------|------|-------|-----------------------|
| مرحله آموزش | | | | |
| ۰/۸۶ | ۰/۷۸۴ | ۲/۴۱ | ۰/۹۱ | GEP ^{W-Haar} |
| ۰/۸۷ | ۰/۷۹۵ | ۲/۳۳ | ۰/۹۲ | GEP ^{W-Coif} |
| ۰/۸۵ | ۰/۷۶ | ۲/۴۵ | ۰/۹۱ | GEP ^{W-Dmey} |
| ۰/۸۳ | ۰/۷۱۷ | ۲/۵۶ | ۰/۹ | GEP |
| مرحله تست | | | | |
| ۰/۸۸ | ۰/۸۲۶ | ۲/۱۳ | ۰/۹۳۳ | GEP ^{W-Haar} |
| ۰/۸۹۹ | ۰/۸۵۳ | ۲ | ۰/۹۴ | GEP ^{W-Coif} |
| ۰/۸۷۹ | ۰/۸۱۸ | ۲/۱۷ | ۰/۹۳ | GEP ^{W-Dmey} |
| ۰/۸۳۹ | ۰/۷۵۳ | ۲/۳۵ | ۰/۹۲ | GEP |



شکل ۲. مقادیر شاخص های خطا در مراحل آموزش و تست



شکل ۳. نمودار راداری شاخص های آماری خطا در مرحله تست

نشان می دهد. این نمودار به صورت مربع هایی هم مرکز می باشد که یکی از کاربردهای اصلی این نمودار، مقایسه مقادیر پیش بینی شده با مقادیر واقعی است. همانطور که در شکل ۳ مشاهده می شود که مقدار RMSE مربوط به روش GEP^{W-Coif} به مرکز نزدیک تر است. درحالی که مقادیر مربوط به شاخص های این روش تمایل بیشتری به خروج از مرکزیت به سمت ۱ را دارد که نشان دهنده

در شکل ۲ نشان داده شده است. بر اساس آن می توان مشاهده نمود که روش GEP^{W-Coif} در هر دو مرحله آموزش و تست برای شاخص RMSE نسبت به سایر روش ها مقدار کمتری دارد. و شاخص های OI, NSE در این روش به عدد یک نزدیکتر است. شکل ۳ نمودار راداری (رادار چارت) مقادیر شاخص خطا در تخمین دمای نقطه شبنم در مرحله تست را

جدول ۴. ارزیابی عملکرد مدل های SVM در مرحله آموزش و تست

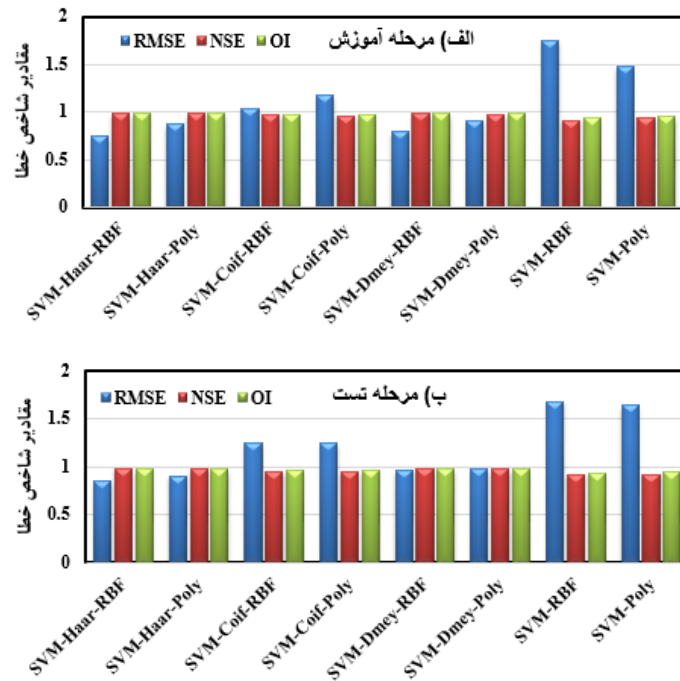
| Model | R | RMSE | NSE | OI |
|-----------------------|-------|------|-------|-------|
| مرحله آموزش | | | | |
| SVM_{RBF}^{W-Haar} | ۰/۹۹ | ۰/۷۴ | ۰/۹۸۴ | ۰/۹۸ |
| SVM_{Poly}^{W-Haar} | ۰/۹۸۸ | ۰/۸۷ | ۰/۹۷۸ | ۰/۹۸ |
| SVM_{RBF}^{W-Coif} | ۰/۹۸ | ۱/۰۴ | ۰/۹۶۸ | ۰/۹۷ |
| SVM_{Poly}^{W-Coif} | ۰/۹۷۹ | ۱/۱۷ | ۰/۹۶ | ۰/۹۶۹ |
| SVM_{RBF}^{W-Dmey} | ۰/۹۹ | ۰/۷۹ | ۰/۹۸۲ | ۰/۹۸ |
| SVM_{Poly}^{W-Dmey} | ۰/۹۸۷ | ۰/۹۱ | ۰/۹۷۶ | ۰/۹۷۹ |
| SVM_{RBF} | ۰/۹۶ | ۱/۷۴ | ۰/۹۱۲ | ۰/۹۴ |
| SVM_{Poly} | ۰/۹۷ | ۱/۴۷ | ۰/۹۳۷ | ۰/۹۵ |
| مرحله تست | | | | |
| SVM_{RBF}^{W-Haar} | ۰/۹۸ | ۰/۸۵ | ۰/۹۸ | ۰/۹۷ |
| SVM_{Poly}^{W-Haar} | ۰/۹۸ | ۰/۹ | ۰/۹۷ | ۰/۹۷ |
| SVM_{RBF}^{W-Coif} | ۰/۹۷ | ۱/۲۴ | ۰/۹۵ | ۰/۹۶ |
| SVM_{Poly}^{W-Coif} | ۰/۹۷ | ۱/۲۴ | ۰/۹۵ | ۰/۹۶ |
| SVM_{RBF}^{W-Dmey} | ۰/۹۸ | ۰/۹۶ | ۰/۹۷ | ۰/۹۷ |
| SVM_{Poly}^{W-Dmey} | ۰/۹۸ | ۰/۹۷ | ۰/۹۷ | ۰/۹۷ |
| SVM_{RBF} | ۰/۹۶ | ۱/۶۸ | ۰/۹۱۶ | ۰/۹۳۷ |
| SVM_{Poly} | ۰/۹۶۵ | ۱/۶۴ | ۰/۹۲ | ۰/۹۴ |

مطلوب بودن عملکرد این روش نسبت به سایر روش هاست.

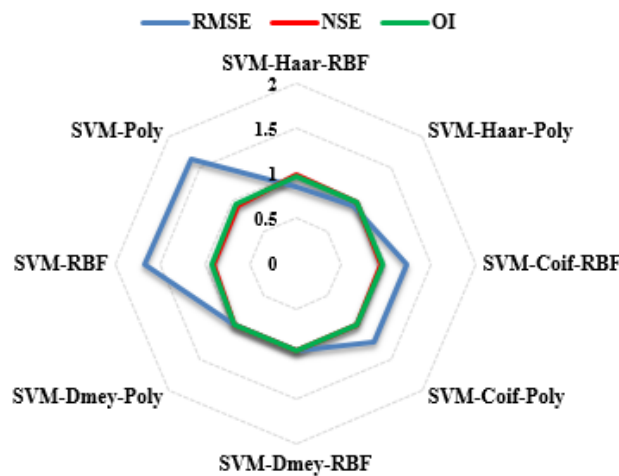
ارزیابی روش ماشین بردار پشتیبان

در این روش از دو کرنل چند جمله ایی Poly و تابع پایه شعاعی RBF جهت مدلسازی دمای نقطه شبنم استفاده شده است. جدول ۴ بیانگر مقدار شاخص های آماری خطا برای مدل های SVM^W و SVM^W در مراحل آموزش و تست می باشد. مطابق جدول ۳ در مرحله آموزش، روش SVM^{W-Haar}_{RBF} با مقادیر NSE=۰/۹۸۴ و RMSE=۰/۷۴ بالاترین دقت را در پیش بینی مقادیر دمای نقطه شبنم را داشت. همچنین، مدل SVM^{W-Dmey}_{RBF}

با ۶/۷۵٪ مقادیر RMSE بیشتر (RMSE=۰/۷۹) و ۰/۲۰٪ مقادیر NSE کمتر (NSE=۰/۹۸۲) در رتبه دوم قرار گرفت. روش SVM^{W-Haar}_{RBF} به ترتیب با کمترین و بیشترین مقادیر ریشه میانگین مربعات خطا و ضریب نش بیشتر (RMSE=۰/۸۵ و NSE=۰/۹۸) به عنوان بهترین مدل در مرحله آموزش انتخاب شد. در حالیکه روش SVM^W_{RBF} با مقادیر RMSE=۱/۶۸ و NSE=۰/۹۱۶ نتوانست عملکرد خوبی را در مقایسه با سایر روش ها از خود نشان دهد. شکل های ۴ و ۵ به ترتیب مقادیر و نمودار راداری شاخص های آماری خطا برای روش های SVM^W و SVM^W به ترتیب در مرحله آموزش و تست را نشان می دهد. مطابق شکل ۴ روش SVM^{W-Haar}_{RBF} به عنوان مطلوب ترین مدل



شکل ۴. مقادیر شاخص های خطا در مراحل آموزش و تست



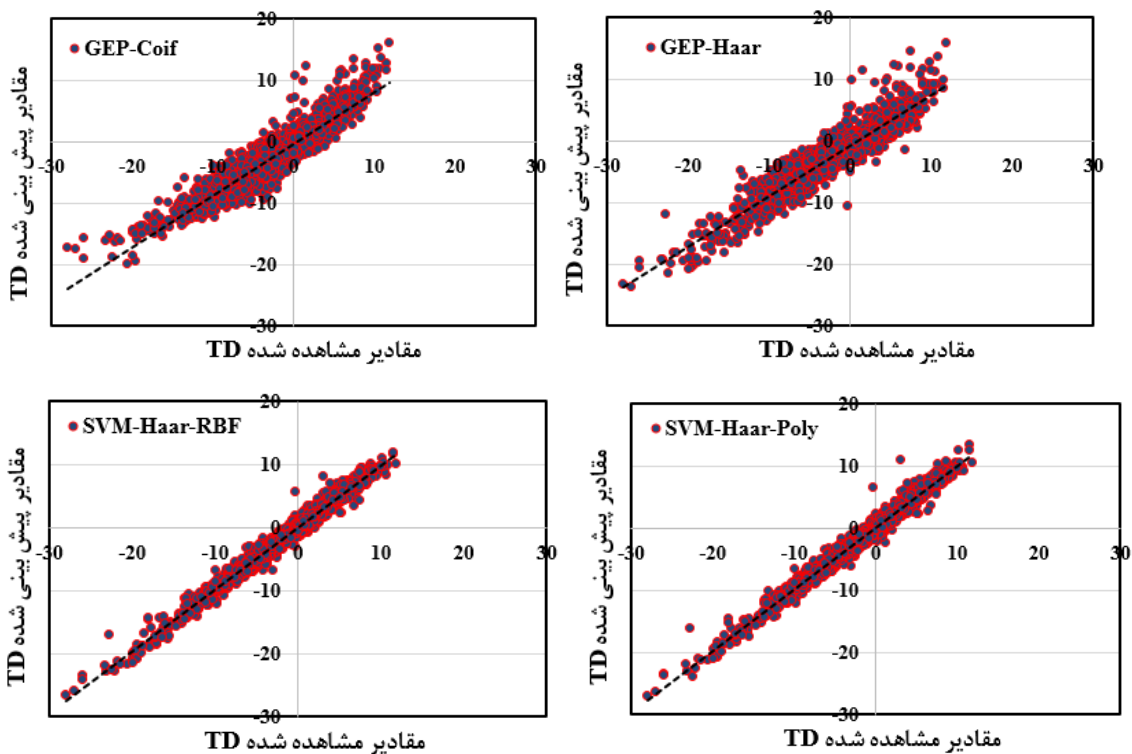
شکل ۵. نمودار راداری شاخص های آماری خطا در مرحله تست

ها در جهت افزایش دقت مدل های SVM و GEP است. در نهایت، جهت مشخص کردن بهترین مدل در تخمین دمای نقطه شبنم، دو مدل برتر از هر روش انتخاب گردید و نتایج در جدول ۵ گزارش شد. با توجه به جدول ۵ می توان نتیجه گرفت که مدل های SVM^W عملکرد بهتری نسبت به مدل های

در بین روش های در نظر گرفته شده انتخاب شد. با توجه به جداول ۳ و ۴ می توان نتیجه گرفت که ترکیب روش تبدیل موجک با برنامه نویسی بیان ژن (GEP^W) و ماشین بردار پشتیبان (SVM^W) از عملکرد بهتری نسبت به روش های منفرد SVM و GEP برخوردار هستند که نشان دهنده تاثیر مثبت روش پیش پردازش کردن داده

جدول ۵. شاخص های آماری برای مدل های برتر در مرحله آموزش و تست

| Model | R | RMSE | NSE | OI |
|-----------------------|------|------|-------|-------|
| مرحله آموزش | | | | |
| SVM_{RBF}^{W-Haar} | ۰/۹۹ | ۰/۷۴ | ۰/۹۸۴ | ۰/۹۸ |
| SVM_{Poly}^{W-Haar} | ۰/۹۸ | ۰/۸۷ | ۰/۹۷۸ | ۰/۹۸ |
| GEP^{W-Haar} | ۰/۹۱ | ۲/۴۱ | ۰/۷۸۴ | ۰/۸۶ |
| GEP^{W-Coif} | ۰/۹۲ | ۲/۳۳ | ۰/۷۹۵ | ۰/۸۷ |
| مرحله تست | | | | |
| SVM_{RBF}^{W-Haar} | ۰/۹۸ | ۰/۸۵ | ۰/۹۸ | ۰/۹۷ |
| SVM_{Poly}^{W-Haar} | ۰/۹۸ | ۰/۹ | ۰/۹۷ | ۰/۹۷ |
| GEP^{W-Haar} | ۰/۹۳ | ۲/۱۳ | ۰/۸۲۶ | ۰/۸۸ |
| GEP^{W-Coif} | ۰/۹۴ | ۲ | ۰/۸۵۳ | ۰/۸۹۹ |



شکل ۶. نمودار پراکندگی مقادیر پیش بینی شده توسط مدل های برتر در برابر مقادیر مشاهده شده در مرحله تست

در مرحله آموزش، SVM_{RBF}^{W-Haar} با بالاترین دقت (۰/۹۸۴) و GEP^{W-Coif} با مقادیر $NSE=۰/۷۹۵$ و $RMSE=۲/۳۳$ نسبت به سایر روش ها مانند

در مرحله آموزش، SVM_{RBF}^{W-Haar} با بالاترین دقت (۰/۹۸۴) و GEP^{W-Coif} با مقادیر $NSE=۰/۷۹۵$ و $RMSE=۲/۳۳$ نسبت به سایر روش ها مانند

نتایج حاصل از بررسی مدل‌ها مشخص شد که مدل SVM_{RBF}^{W-Haar} توانست در مرحله آموزش با معیارهای خطا $NSE=0/984$ و $R=0/99$ توانست به خوبی مقادیر TD را تخمین بزند. در مرحله تست، مدل SVM_{RBF}^{W-Haar} نیز عملکرد قابل قبولی را از خود نشان داد و با کمترین مقدار خطا ($RMSE=0/85$) در رتبه اول قرار بگیرد. همچنین لازم بذکرست که مدل‌های برتر GEP^W نتوانستند به نسبت مدل‌های برتر SVM^W مقادیر TD را تخمین بزنند. بطور مثال، مدل GEP^{W-Coif} با مقادیر $OI=0/899$ و $RMSE=2$ که بهترین عملکرد در گروه خود در مرحله تست را داشت، در مقایسه با SVM_{Poly}^{W-Haar} ($OI=0/97$) و $RMSE=0/9$ دارای دقت پایین تری بود.

توانست در رتبه اول قرار بگیرد. علاوه بر این، شاخص‌های آماری خطا نشان دادند که مقادیر پیش‌بینی شده توسط مدل SVM_{RBF}^{W-Haar} به مقادیر مشاهداتی نزدیک‌تر از سایر مدل‌های در نظر گرفته شده است.

شکل ۶ نمودار پراکندگی مقادیر پیش‌بینی شده توسط برترین مدل‌ها از هر روش در برابر مقادیر مشاهداتی در مرحله تست را نشان می‌دهد. براساس آن می‌توان نتیجه گرفت که پراکندگی مقادیر پیش‌بینی شده TD توسط مدل‌های SVM_{Poly}^{W-Haar} و SVM_{RBF}^{W-Haar} کمتر از مدل‌های GEP^W و GEP^{W-Coif} است و می‌توان مشاهده نمود که مقادیر TD پیش‌بینی شده توسط مدل SVM_{RBF}^{W-Haar} به خط ایده‌آل نزدیکتر است.

مراجع

- Abhinav, R., Pindoriya, N. M., Wu, J., Long, C. 2017. Short-term wind power forecasting using wavelet-based neural network. *Energy Procedia*, 142, 455-460.
- Alizadeh, M. J., Kavianpour, M. R., Danesh, M., Adolf, J., Shamshirband, S., Chau, K. W. 2018. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. *Engineering. Applications of Computational Fluid Mechanics*, 12(1), 810-823.
- Al-Shammari, E. T., Mohammadi, K., Keivani, A., Ab Hamid, S. H., Akib, S., Shamshirband, S., Petković, D. 2016. Prediction of daily dewpoint temperature using a model combining the support vector machine with firefly algorithm. *Journal of Irrigation and Drainage Engineering*, 142(5), 04016013.
- Amirmojahedi, M., Mohammadi, K., Shamshirband, S., Danesh, A. S., Mostafaeipour, A., Kamsin, A. 2016. A hybrid computational intelligence method for predicting dew point temperature. *Environmental Earth Sciences*, 75(5), 415.
- Attar, N. F., Khalili, K., Behmanesh, J., Khanmohammadi, N. 2018. On the reliability of soft computing methods in the estimation of dew point temperature: The case of arid regions of Iran. *Computers and electronics in agriculture*, 153, 334-346.
- Baghban, A., Bahadori, M., Rozyn, J., Lee, M., Abbas, A., Bahadori, A., Rahimali, A. 2016. Estimation of air dew point temperature using computational intelligence schemes. *Applied thermal engineering*, 93, 1043-1052.
- Deka, P. C., Patil, A. P., Yeswanth Kumar, P., Naganna, S. R. 2018. Estimation of dew point temperature using SVM and ELM for humid and semi-arid regions of India. *ISH Journal of Hydraulic Engineering*, 24(2), 190-197.
- Deo, R. C., Samui, P., Kim, D. 2016. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stochastic Environmental Research and Risk Assessment*, 30(6), 1769-1784.

جمع‌بندی

دمای نقطه شبنم به عنوان یکی از مهمترین پدیده‌های هیدرولوژیکی، در تولید محصولات کشاورزی، زندگی انسان و حیوانات، تغییرات آب و هوایی و پیش‌بینی بعضی از پدیده‌های هیدرولوژیکی نقش اساسی و کلیدی را ایفا می‌کند. از طرفی با توجه به کمبود منابع آب‌های زیر زمینی رطوبت موجود در هوا می‌تواند موثر واقع شود. از این رو تخمین دقیق دمای نقطه شبنم از اهمیت فراوانی برخوردار هستند. برای پیش‌بینی دمای نقطه شبنم از متغیرهای دمای حداکثر و حداقل، سرعت باد، رطوبت نسبی و ساعت تابش خورشید که به صورت روزانه در مدت ۱۹ سال اندازه‌گیری شده‌اند، استفاده شده است. جهت تحقق این هدف از روش برنامه‌نویسی بیان ژن و ماشین بردار پشتیبان و ترکیب آنها با یکی از روش‌های پیش‌پردازش داده‌ها بنام تبدیل موجک استفاده شد. برای ارزیابی مدل‌ها از شاخص‌های آماری $RMSE$ ، R ، NSE و OI استفاده شد. در ابتدا هر کدام از مدل‌های SVM^W و GEP^W با مدل‌های SVM و GEP بصورت جداگانه مورد بررسی قرار گرفتند که مشخص شد موجک باعث بهبود دقت مدل شده است. در انتها، بمنظور انتخاب بهترین مدل، از هر قسمت دو مدل برتر را انتخاب گردید و مورد ارزیابی قرار گرفتند. براساس

- 3), 1-5.
- Li, L. L., Zhao, X., Tseng, M. L., & Tan, R. R. 2020. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production*, 242, 118447.
- Mehdizadeh, S., Behmanesh, J., Khalili, K. 2017. Application of gene expression programming to predict daily dew point temperature. *Applied Thermal Engineering*, 112, 1097-1107.
- Mehdizadeh, S., Ahmadi, F., Mehr, A. D., Safari, M. J. S. 2020. Drought modeling using classic time series and hybrid wavelet-gene expression programming models. *Journal of Hydrology*, 587, 125017.
- Mohammadi, k., Shamshirband, S., Petkovic, D., Lip Yee, P. 2016. Using ANFIS for selection of more relevant parameters to predict dew point temperature. *Thermal Engineering*, 311-319.
- Mohammadi, K., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., Gocic, M. 2015. Extreme learning machine based prediction of daily dew point temperature. *Computers and Electronics in Agriculture*, 117, 214-225.
- Montaseri, M., Ghavidel, S. Z. Z., Sanikhani, H. 2018. Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques. *Stochastic Environmental Research and Risk Assessment*, 32(8), 2253-2273.
- Naganna, S. R., Deka, P. C., Ghorbani, M. A., Biazar, S. M., Al-Ansari, N., Yaseen, Z. M. 2019. Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water*, 11(4), 742.
- Noori, R., Abdoli, M. A., Ameri, A., Jalili-Ghazizade, M. 2008. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad. *Environmental Progress and Sustainable Energy*, 28 (2), 249-258.
- Qasem, S. N., Samadianfard, S., Sadri Nahand, H., Mosavi, A., Shamshirband, S., Chau, K. W. 2019. Estimating daily dew point temperature using machine learning algorithms. *Water*, 11(3), 582.
- Rezaie-balf, M., Naganna, S. R., Ghaemi, A., Deka, P. C. 2017. Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of Hydrology*, 553, 356-373.
- Shiri, J. 2019. Prediction vs. estimation of dewpoint temperature: assessing GEP, MARS and RF models. *Hydrology Research*, 50(2), 633-643.
- Yu, H., Ming, L. J., Sumei, R., Shuping, Z. 2020. A hybrid model for financial time series forecasting—integration of EWT, ARIMA with the improved ABC optimized ELM. *IEEE Access*, 8, 84501-84518.
- Zounemat-Kermani, M. 2012. Hourly predictive Levenberg–Marquardt ANN and multi linear regression models for predicting of dew point temperature. *Meteorology and Atmospheric Physics*, 117(3-4), 181-192.
- Zamani, R., Ahmadi, F., Radmanesh, F. 2015. Comparison of the Gene Expression Programming, Nonlinear Time Series and Artificial Neural Network in Estimating the River Daily Flow (Case Study: The Karun River). *Water and Soil*, 28(6), 1172-1182.
- Deo, R. C., Tiwari, M. K., Adamowski, J. F., Quilty, J. M. 2017. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stochastic environmental research and risk assessment*, 31(5), 1211-1240.
- Dong, J., Wu, L., Liu, X., Li, Z., Gao, Y., Zhang, Y., Yang, Q. 2020. Estimation of daily dew point temperature by using bat algorithm optimization based extreme learning machine. *Applied Thermal Engineering*, 165, 114569.
- Gandomi AH, Alavi AH, Sahab MG, Arjmandi P. 2010. Formulation of elastic modulus of concrete using linear genetic programming. *Journal of Mechanical Science and Technology*, 24(6), 1273-1278.
- Feng, Z. K., Niu, W. J., Tang, Z. Y., Jiang, Z. Q., Xu, Y., Liu, Y., Zhang, H. R. 2020. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *Journal of Hydrology*, 583, 124627.
- Ghaemi, A., Rezaie-Balf, M., Adamowski, J., Kisi, O., Quilty, J. 2019. On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. *Agricultural and Forest Meteorology*, 278, 107647.
- Ghaemi, A., Zhian, T., Pirzadeh, B., Monfared, S. H., Mosavi, A. 2021. Reliability-based design and implementation of crow search algorithm for longitudinal dispersion coefficient estimation in rivers. *Environmental Science and Pollution Research*, 1-20.
- Graf, R., Zhu, S., Sivakumar, B. 2019. Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *Journal of Hydrology*, 578, 124115.
- Kaboli, S. H. A., Fallahpour, A., Selvaraj, J., Rahim, N. A. 2017. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. *Energy*, 126, 144-164
- Kazemi, M. H., Majnooni-Heris, A., Kisi, O., Shiri, J. 2021. Generalized gene expression programming models for estimating reference evapotranspiration through cross-station assessment and exogenous data supply. *Environmental Science and Pollution Research*, 28(6), 6520-6532.
- Kim, S., Seo, Y., Rezaie-Balf, M., Kisi, O., Ghorbani, M. A., Singh, V. P. 2019. Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: peninsula vs continent. *Theoretical and Applied Climatology*, 137(1-2), 693-712.
- Kisi, O., Parmar, K. S. 2015. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology*, 534, 104-112.
- Kisi, O., Shiri, J., & Tombul, M. 2013. Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, 51, 108-117.
- Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., ... & Bui, D. T. 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*, 627, 744-755.
- Lashkaripour, G. R., Zivdar, M. 2005. Desalination of brackish groundwater in Zahedan city in Iran. *Desalination*, 177(1-

Evaluating the effect of data preprocessing methods on the performance of soft computing methods in estimation of dew point temperature

Mohammad Teimuri¹, Alireza Ghaemi², Gholamreza Azizyan³, Seyed Arman Hashemi Monfared^{4*}

¹ MSc Student of Civil Engineering, Faculty of water resource management, University of Sistan and Baluchestan, Zahedan, Iran

² PhD Student of Civil Engineering, Faculty of water resource management, University of Sistan and Baluchestan, Zahedan, Iran

³ Associate Professor of Civil Engineering, Faculty of water resource management, University of Sistan and Baluchestan, Zahedan, Iran

⁴ Associate Professor of Civil Engineering, Faculty of water resource management, University of Sistan and Baluchestan, Zahedan, Iran

*Corresponding Author Email: hashemi@eng.usb.ac.ir

Received: 08 March 2020, accepted: 13 June 2020

ABSTRACT

The temperature which the amount of saturated vapor pressure is equal to the actual pressure and due to decreasing it, the existence moisture of air also turns into a liquid is called dew point temperature. Additionally, owing to the shortage of water in whole world, water supplying to solve the initial requirements of human has become a global challenge. Due to the limited number of dew point measuring stations, employing the various approaches with acceptable accuracy is required. Recently, due to the significant ability of different soft computing (SC) methods in communizing between inputs and outputs, they have been considered by researchers to solve a wide range of problems. Therefore, in the present research, the performance of two SC methods namely support vector machine (SVM) and gene expression programming (GEP) for prediction of dew point temperature was investigated. Additionally, wavelet transform (WT) as a data preprocessing method was employed to improve the accuracy of models. Evaluation of models performance based on the statistical error benchmarks indicated that hybrid models outperformed the standalone SVM and GEP approaches. Moreover, the results illustrated that the accuracy of the hybrid model obtained by integration of WT and GEP (GEPW) was less than its integration with support vector machine (SVMW). Finally, among the proposed methods, model was able to estimate the dew point temperature values with the highest accuracy with respect to NSE (0.98) and RMSE (0.85) and determined as the best model so as to estimate the dew point temperature.

Keywords: Dew point temperature, support vector machine, gene expression programming, wavelet transform

HOW TO CITE THIS ARTICLE

Teimuri, M.; Ghaemi, A.; Azizyan G.; Hashemi Monfared SR. (2020). Evaluating the effect of data preprocessing methods on the performance of soft computing methods in estimation of dew point temperature. *J. Meteorol. Atmos. Sci.*, 3(2): 175-187.

COPYRIGHTS

Copyright for this article is retained by the author(s), with publication rights granted to the JMAS Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

