

ارزیابی مدل‌های هوشمند در برآورد میزان مواد جامد محلول در آب رودخانه کشکان، استان لرستان

الهام رضایی^۱، بابک شاهی نژاد^{۲*} و حجت‌اله یونسی^۳

^۱ دانشجوی کارشناسی ارشد سازه‌های آبی، دانشکده کشاورزی، دانشگاه لرستان، ^۲ استادیار، دانشکده کشاورزی، دانشگاه لرستان

تاریخ پذیرش: ۱۳۹۷/۰۶/۲۷

تاریخ دریافت: ۱۳۹۷/۰۵/۰۷

چکیده

یکی از موضوعات مهم در بحث کیفی رودخانه‌ها پیش‌بینی میزان مواد جامد محلول در آب می‌باشد. در این پژوهش، عملکرد مدل‌های هوشمند ماشین بردار پشتیبان با توابع هسته مختلف، برنامه‌ریزی بیان ژن و شبکه بیزین برای پیش‌بینی میزان مواد جامد محلول در آب رودخانه کشکان مورد بررسی قرار گرفت. بدین‌منظور، از داده‌های کیفی ایستگاه پلدختر واقع در استان لرستان، شامل هیدروژن کربنات، کلرید، سولفات، منیزیم، کلسیم، سدیم، هدایت الکتریکی، دبی جریان و pH برای پیش‌بینی میزان مواد جامد محلول در آب طی دوره آماری ۱۳۹۵-۱۳۷۰ در مقیاس زمانی ماهانه استفاده شد. برای صحت‌سنجی مدل‌ها از معیارهای ضریب تبیین، ضریب نش-ساتکلیف، ریشه میانگین مربعات خطا و اریب استفاده شد. نتایج به‌دست آمده نشان داد که در هر سه مدل مذکور، ساختارهای ترکیبی از دقت قابل قبولی برخوردار هستند. همچنین، بر اساس معیارهای ارزیابی مشخص شد که مدل ماشین بردار پشتیبان با هسته توابع پایه شعاعی بیشترین دقت ۰/۹۸۲، کمترین میزان ریشه مربعات خطا (mg lit^{-1}) ۰/۰۳۲، کمترین اریب ۰/۰۰۱ و بیشترین ضریب نش-ساتکلیف ۰/۹۶۳ را نسبت به سایر مدل‌ها داشته است.

کلمات کلیدی: برنامه‌ریزی بیان ژن، پیش‌بینی، توابع هسته مختلف، شبکه بیزین، ماشین بردار پشتیبان

مقدمه

افزایش جمعیت، موجب افزایش نیازها و به‌دنبال آن استفاده بیش از حد از منابع آبی، دخل و تصرف غیر طبیعی و تغییر شرایط کمی و کیفی آب رودخانه‌ها شده است. این امر منجر به آلودگی رودخانه‌ها شده که یکی از معضلات دنیای امروز محسوب می‌شود (Hosseini و همکاران، ۲۰۱۳). با شناخت کافی از متغیرهای کیفی می‌توان از یک مدل مناسب برای شبیه‌سازی کیفی و استفاده از نتایج آن در انجام اقدامات مدیریتی استفاده نمود. از آنجایی که عامل‌های زیادی بر پارامترهای کیفیت آب مؤثر

رودخانه‌ها از منابع مهم آب‌های سطحی می‌باشند. کیفیت آب‌های سطحی تحت تأثیر فرآیندهای طبیعی شامل شرایط اقلیمی، وجود سازندهای حساس و انتقال رسوب و فعالیت‌های انسانی شامل توسعه و گسترش شهرنشینی، تغییر کاربری اراضی و تخلیه پساب‌های صنعتی و کشاورزی قرار دارند (Muangthong، ۲۰۱۵). کیفیت آب رودخانه‌ها نقش مهمی در سلامت انسان و توسعه اجتماعی و اقتصادی دارد (Xu و همکاران، ۲۰۱۶). پیشرفت‌های صنعتی و

در سال‌های اخیر استفاده از مدل‌های هوشمند در مطالعه و پیش‌بینی متغیرهای کیفی آب رودخانه‌ها مورد توجه پژوهشگران مختلف قرار گرفته است. برای مثال Eskandari و همکاران (۲۰۱۲) به توسعه مدلی بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای تخمین به‌هنگام اکسیژن‌خواهی بیوشیمیایی^۵ در رودخانه سفیدرود پرداختند. نتایج به‌دست آمده حاکی از عملکرد برتر مدل ماشین بردار پشتیبان نسبت به شبکه عصبی مصنوعی بود. در پژوهشی Zeinalzadeh و Zamanzad Ghavidel (۲۰۱۶) عملکرد مدل‌های شبکه عصبی مصنوعی، سیستم استنتاجی عصبی- فازی و برنامه‌ریزی بیان ژن در مدل‌بندی خصوصیات کیفی آب و برآورد میزان مواد جامد محلول در رودخانه بالادست سد بوکان واقع در حوزه آبخیز زرینه‌رود مورد بررسی قرار دادند. با وجود عملکرد قابل قبول هر سه روش، مدل برنامه‌ریزی بیان ژن علاوه بر عملکرد بالا، روابط ریاضی برای برآورد مقدار املاح محلول در رودخانه‌های بالادست سد را نیز ارائه داده و نسبت به سایر مدل‌های مذکور برتری داشته است. در مطالعه‌ای دیگر Montaseri و Zamanzad Ghavidel (۲۰۱۶) عملکرد شبکه عصبی مصنوعی و استنتاجی عصبی- فازی را برای تخمین پارامترهای کیفی آب رودخانه زرینه‌رود در دو دوره کم‌آبی و پرآبی مورد ارزیابی قرار دادند. نتایج نشان داد، عملکرد روش عصبی-فازی تطبیقی در هر دو دوره مطالعاتی بهتر از شبکه‌های عصبی مصنوعی می‌باشد. همچنین، Asadollahfardi و همکاران (۲۰۱۶) شبکه عصبی پایه شعاعی (RBF) و شبکه عصبی تاخیر زمانی (TDNN^۱) را برای پیش‌بینی مواد جامد محلول در رودخانه زاینده‌رود ایران استفاده نمودند. مقایسه نتایج حاصل از دو مدل نشان داد که مدل پایه شعاعی عملکرد مناسبی برای پیش‌بینی مواد جامد محلول رودخانه داشته و نسبت به مدل تاخیر زمانی از دقت بالاتری برخوردار است. سپس Sattari و همکاران (۲۰۱۶) به‌منظور برآورد دو پارامتر هدایت الکتریکی و مواد جامد محلول رودخانه

هستند و یک رابطه غیر خطی پیچیده بین آن‌ها برقرار است، بنابراین، روش‌های سنتی پردازش داده نمی‌توانند مسئله مدیریت کیفی منابع آب را به‌خوبی حل کنند (Singh و همکاران، ۲۰۰۹). مدل‌های هوشمند به‌دلیل دقت و سرعت بالای پردازش و هزینه کم نسبت به مدل‌های دیگر، گسترش زیادی یافته‌اند که برای شبیه‌سازی و پیش‌بینی متغیرهای منابع آب مناسب می‌باشند (Rezaei و همکاران، ۲۰۱۴). از جمله مدل‌های هوشمند می‌توان به مدل شبکه بیزین (BN^۱)، برنامه‌ریزی بیان ژن (GEP^۲) و ماشین بردار پشتیبان (SVM^۳) اشاره نمود. شبکه بیزین جزئی از خانواده مدل‌های گرافیکی احتمالی است که مجموعه‌ای از متغیرها و احتمالات مربوط به هر کدام را نشان می‌دهد. ساختار یک شبکه بیزین در واقع یک نمایش نموداری از اثرات متقابل متغیرهایی است که باید مدل شوند (Cruz-Ramírez و همکاران، ۲۰۰۷). برنامه‌ریزی بیان ژن که عضوی از خانواده الگوریتم تکاملی است، یک تکنیک برنامه‌ریزی خودکار می‌باشد که راه‌حل مسئله را با استفاده از برنامه‌ریزی کامپیوتری ارائه می‌دهد (Salehi و Ghorbani، ۲۰۱۱). بر عکس، بسیاری از مدل‌های پایه آماری کیفیت آب که با فرض رابطه خطی بین مقادیر واقعی و تخمینی عمل می‌نمایند، روش مذکور که برگرفته از طبیعت می‌باشد، قادر به ارائه رابطه غیر خطی حاکم بر مسئله می‌باشد. ماشین بردار پشتیبان نیز یکی از روش‌های یادگیری نظارت شده بر مبنای تئوری بهینه‌سازی مقید^۴ است که اصل حداقل‌سازی خطای ساختاری را به‌کار گرفته و منجر به یک جواب بهینه کلی می‌شود (Vapnik، ۱۹۹۸). فرایند بهینه‌سازی تابع هدف در فضای یک یا چند محدودیت (قید) را بهینه‌سازی مقید می‌گویند. تابع هدف ممکن است یک تابع هزینه باشد که باید به حداقل برسد، یا یک تابع مفید باشد، که باید حداکثر شود. در مدل ماشین بردار پشتیبان هدف کمینه کردن خطای ساختاری می‌باشد.

^۱ Bayesian Network^۲ Gene Expression Programming^۳ Support Vector Machine^۴ Constraint Optimization^۵ Biochemical Oxygen Demand (BOD5)^۶ Time-Delay Neural Network

رودخانه تیره واقع در جنوب غربی ایران از شبکه عصبی مصنوعی، روش دسته‌بندی داده‌ها و ماشین بردار پشتیبان استفاده نمودند. نتایج به‌دست آمده حاکی از عملکرد قابل قبول هر سه مدل جهت پیش‌بینی پارامترهای کیفی آب بوده، اما مدل SVM با تابع هسته RBF عملکرد بهتری ارائه نموده است. در پژوهشی Solgi و همکاران (۲۰۱۸) کاربرد مدل ماشین بردار پشتیبان و مدل ترکیبی ماشین بردار پشتیبان-موجک به‌منظور مدل‌سازی و پیش‌بینی اکسیژن مورد نیاز بیولوژیکی^۵ در رودخانه کارون مورد ارزیابی قرار دادند. نتایج آن‌ها حاکی از آن بود که اعمال تبدیل موجک به مدل SVM باعث بهبود نتایج می‌شود.

مروری بر مطالعات گذشته اهمیت استفاده از مدل‌های هوشمند را در پیش‌بینی و برآورد پارامترهای کیفی رودخانه نشان می‌دهد. با توجه به وجود روابط غیر خطی بین پارامترهای کیفی رودخانه، استفاده از روش‌های هوشمند در پیش‌بینی و برآورد پارامترهای یاد شده امری ضروری می‌باشد. از طرفی، مقدار مواد جامد محلول در آب به‌عنوان یک عامل مهم در بحث رسوب و کیفیت آب به‌ویژه در رودخانه‌های بزرگ مطرح است و می‌تواند در تصمیم‌گیری‌های مدیریتی تاثیرگذار باشد. به همین علت در این تحقیق سعی شد، عملکرد سه روش ماشین بردار پشتیبان با توابع هسته مختلف، برنامه‌ریزی بیان ژن و شبکه بیزین در برآورد کل مواد جامد محلول در رودخانه کشکان به‌عنوان یکی از مهمترین رودخانه‌های استان لرستان و تأمین‌کننده آب بخش‌های مختلف و نواحی مجاور خود از لحاظ کشاورزی و شرب، مورد ارزیابی و بررسی قرار گرفته و در نهایت مدل برتر برای مدل‌سازی کیفی آب این رودخانه و استفاده از نتایج آن در اقدامات مدیریتی ارائه شود.

مواد و روش

منطقه مورد پژوهش و داده‌ها: حوزه آبخیز رودخانه کشکان در ناحیه جنوب غربی ایران و در بین طول‌های جغرافیایی "۳۴' ۳۱' ۴۷" تا "۶' ۱۲' ۴۸"

لیقوان چای واقع در آذربایجان شرقی با استفاده از الگوریتم k- نزدیک‌ترین همسایه و رگرسیون بردار پشتیبان به این نتیجه رسیدند که هر دو الگوریتم به‌طور دقیق مواد جامد محلول و هدایت الکتریکی را برآورد می‌کنند، اما مدل رگرسیون بردار پشتیبان به‌دلیل عملکرد و دقت بهتر، توصیه می‌شود. در مطالعه‌ای Kanda و همکاران (۲۰۱۶) از شبکه‌های عصبی مصنوعی برای پیش‌بینی اکسیژن محلول رودخانه نزویا واقع در غرب کنیا استفاده کردند. نتایج گویای توانایی بالای شبکه عصبی مصنوعی بود. در پژوهشی Liyanage و Yamada (۲۰۱۶) نیز به طراحی دو مدل طبقه‌بندی کیفیت آب با استفاده از شبکه عصبی مصنوعی و شبکه بیزین پرداختند. نتایج نشان داد، مدل بیزین دقت بالاتری نسبت به مدل شبکه عصبی دارد. در بررسی Dehghani و Ghorbani (۲۰۱۷) عملکرد مدل‌های شبکه عصبی بیزین، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن برای تحلیل میزان مواد جامد محلول در آب رودخانه بالخلوچای واقع در استان اردبیل را مورد بررسی قرار دادند. نتایج حاصله نشان داد که مدل شبکه عصبی بیزین از دقت بالاتری برخوردار است. در پژوهشی Bozorg-Haddad و همکاران (۲۰۱۷) در مدل‌سازی پارامترهای کیفیت آب سفیدرود ایران با استفاده از الگوریتم کمترین مربعات پشتیبانی رگرسیون واریانس و برنامه‌نویسی ژنتیکی به این نتیجه رسیدند که الگوریتم GA-LSSVR دقت بیشتری برای مدل‌سازی پارامترهای کیفیت آب نسبت به برنامه‌ریزی ژنتیک دارد. در پژوهشی Khaled و همکاران (۲۰۱۸) به‌منظور مدل‌سازی اکسیژن‌خواهی بیوشیمیایی^۱ از سیستم استنتاج عصبی-فازی تطبیقی^۲ از دو روش افزایش شبکه^۳ و دسته‌بندی تفریقی^۴ استفاده نمودند. نتایج ارائه شده بیان‌گر عملکرد و دقت بهتر روش افزایش شبکه ANFIS-SC در برآورد BOD5 نسبت به ANFIS-GP بوده است. در پژوهشی Hagiabi و همکاران (۲۰۱۸) به‌منظور پیش‌بینی کیفیت آب

¹ Biochemical Oxygen Demand (BOD5)

² Adaptive Neuro Fuzzy Inference System (ANFIS)

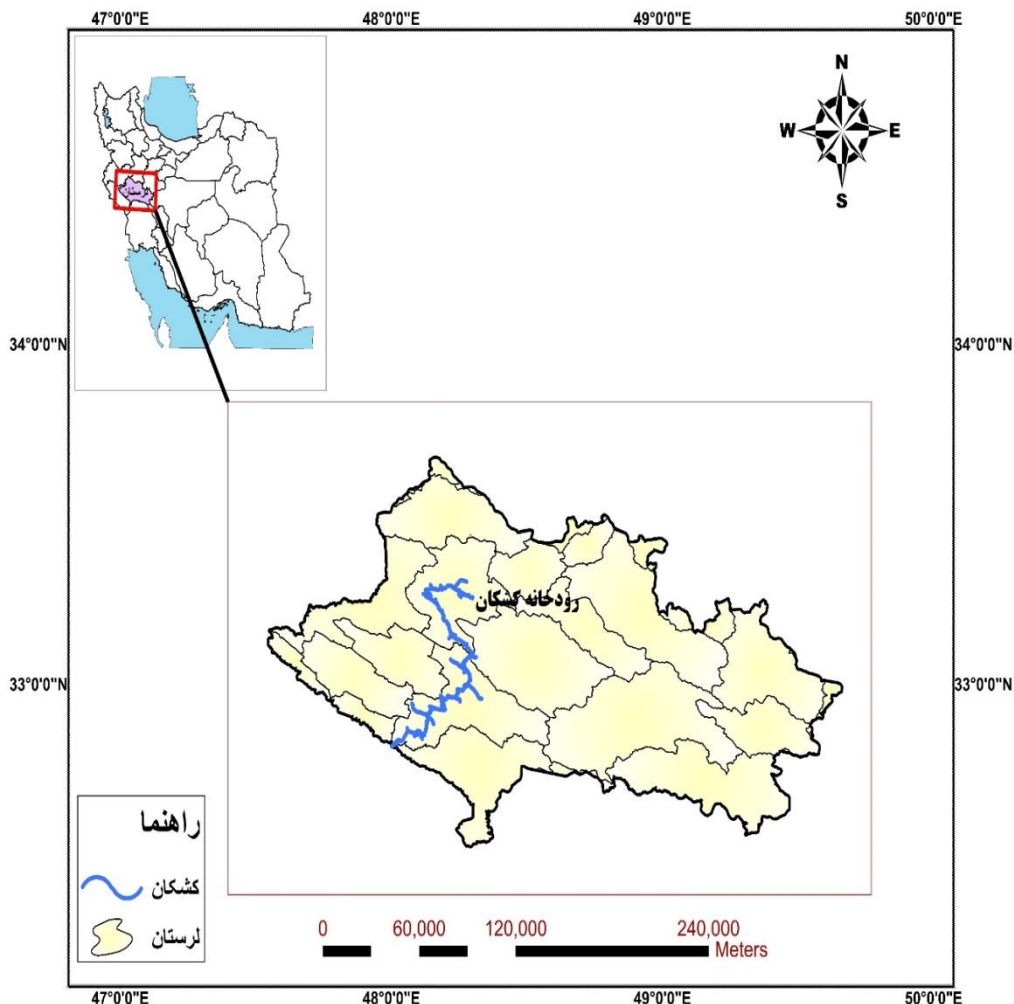
³ Grid Partitioning (ANFIS-GP)

⁴ Subtractive Clustering (ANFIS-SC)

⁵ Biochemical Oxygen Demand (BOD)

را نشان می‌دهد. در این پژوهش، از متغیرهای کمی و کیفی مربوط به ایستگاه رودخانه کشکان در مقیاس زمانی ماهانه، شامل متغیرهای هیدروژن کربنات (HCO_3)، کلرید (Cl)، سولفات (SO_4)، منیزیم (Mg)، کلسیم (Ca)، سدیم (Na) و هدایت الکتریکی (EC) همگی برحسب میلی‌گرم بر لیتر و دبی جریان (Q) برحسب لیتر بر ثانیه و نیز pH به‌عنوان ورودی مدل و مواد جامد محلول (TDS) بر حسب میلی‌گرم بر لیتر به‌عنوان خروجی مدل، طی دوره آماری ۱۳۷۰-۱۳۹۵ استفاده شد.

شرقی و عرض‌های جغرافیایی $33^\circ 5' 45''$ تا $41'$ $44'$ شمالی واقع شده است. این حوضه بخش مهمی از سرشاخه‌های پر آب رودخانه کرخه را تشکیل می‌دهد و حدود یک‌سوم از خاک استان لرستان را در بر گرفته و در تقسیم‌بندی هیدرولوژی ایران، جزئی از حوزه آبخیز خلیج فارس به شمار می‌رود. رودخانه مورد بررسی، بخشی از حوزه آبخیز کشکان در زیرحوضه کشکان پلدختر است که از محل ایستگاه هیدرومتری کشکان-پلدختر شروع و با عبور از شهر پلدختر تا محل پل گاومیشان در استان ایلام ادامه دارد. شکل ۱، موقعیت جغرافیایی منطقه مورد مطالعه



شکل ۱ - موقعیت جغرافیایی منطقه مورد مطالعه

الگوریتم ژنتیک (GA) است که در سال ۱۹۹۹ توسط Ferreira بر اساس تئوری Darwin ابداع شد.

برنامه‌ریزی بیان ژن: برنامه‌ریزی بیان ژن (GEP)، توسعه‌یافته روش‌های برنامه‌ریزی ژنتیک (GP) و

احتمالات وابسته یا قانون بیز می‌باشد. رابطه زیر قانون بیز را نشان می‌دهد.

$$P(b|a) = P(a|b) \times \frac{P(b)}{P(a)} \quad (1)$$

که در آن، $P(a)$ احتمال وقوع پیشامد a ، $P(b)$ احتمال وقوع پیشامد b ، $P(b|a)$ احتمال وقوع پیشامد b به شرطی که a اتفاق افتاده باشد و $P(a|b)$ احتمال وقوع پیشامد a به شرطی که b اتفاق افتاده باشد، هستند.

هر شبکه بیزین از سه جزء اصلی تشکیل شده است: ۱- گره‌ها: که در واقع متغیرهای سیستم مدیریتی هستند. ۲- بندها: روابط سببی بین متغیرها به وسیله بندها به وجود می‌آیند. ۳- احتمالات: هر کدام برای متغیر حالت تصمیمی را مشخص کرده و از طریق متغیرهایی که مستقیماً آن را تحت تأثیر قرار می‌دهند (والدین آن)، به آن نسبت داده می‌شوند (Ahmadi و همکاران، ۲۰۱۶). بیزین یک ساختار گرافیکی می‌باشد که متغیرها به صورت گره و روابط علت و معلولی آن‌ها با استفاده از یک سری پیکان موسوم به Edge مشخص می‌شوند. در حالت کلی، گره‌ها یا والد هستند یا ولد (فرزند). یک گره فرزند می‌تواند از سوی چندین والد تولید شده باشد. گره‌هایی که قبل از آن‌ها گره دیگری در گراف وجود دارد، با توزیع احتمال شرطی تعریف می‌شوند، در غیر این صورت با احتمال آغازین (اولیه) بیان می‌شوند. پایین‌ترین بخش در شبکه بیزین از طریق قانون احتمال کل به دست می‌آید. احتمالات مرتبط با بخش‌های بالایی شبکه بر اساس قانون بیز می‌باشند. در حالت کلی، اگر گره x دارای والد نباشد گره غیرمشروط و در غیر این صورت گره مشروط می‌باشد (Kuikka و Varis، ۱۹۹۷).

در مورد ساختار شبکه بیزین اگر مجموعه‌ای شامل n متغیر به صورت $\{X_1, X_2, X_3, \dots, X_n\}$ وجود داشته باشد که هر X_i یک متغیر تصادفی باشد و گره‌های والد این متغیر با $parent(X_i)$ نشان داده شوند، احتمال هر رویداد مانند $X=(x_1, x_2, x_3, \dots, x_n)$ به صورت زیر محاسبه می‌شود.

$$P(x_1 + x_2 + \dots + x_n) = \prod P(x_i | parents(x_i)) \quad (2)$$

به این ترتیب که جمعیتی به صورت انتخابی در جهت تکامل، جمعیت نامناسب را رها کرده و فرزندان اصلاح شده ایجاد می‌کنند. در این روش، در ابتدای فرایند هیچ‌گونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مؤلفه‌های آن می‌باشد. در روش برنامه‌ریزی بیان ژن، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با الگوریتم ژنتیک و ساختارهای شاخه‌ای با اندازه و شکل‌های متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک ترکیب می‌شوند. به طور خلاصه می‌توان بیان کرد که در این روش، فنوتیپ و ژنوتیپ از هم جدا شده و سیستم قادر خواهد بود از تمام مزایای تکاملی بهره گیرد. با وجود این که فنوتیپ در GEP مشابه ساختار شاخه‌ای GP می‌باشد، اما ساختار شاخه‌ای در GEP که بیان درختی نیز نامیده می‌شود، بیان‌گر تمام ژنوم‌های مستقل است. در مجموع، در GEP به‌سازی در یک ساختار خطی اتفاق افتاده و سپس به صورت ساختار درختی بیان می‌شود. این امر سبب خواهد شد که تنها ژنوم اصلاح شده به نسل بعد انتقال یابد و نیازی به ساختارهای سنگین تکثیر و جهش نباشد. برنامه‌ریزی بیان ژن برخلاف الگوریتم ژنتیک روی ساختار درختی رابطه‌ها به جای سلسله ارقام دودویی عمل می‌کند که این ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در رابطه‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند (Ferreira، ۲۰۰۱).

در این پژوهش از عملگرهای F1 شامل چهار عملگر $+$ ، $-$ ، \times ، $/$ و عملگرهای F2 شامل عملگرهای $+$ ، $-$ ، \times ، $/$ ، $\sqrt{\quad}$ ، $\sqrt[3]{\quad}$ ، Exp ، Ln ، 2^{\quad} ، 3^{\quad} ، Sin ، Cos ، Atan استفاده شد و این عملگرها در نرم‌افزار Genxprotools 5 به راحتی قابل دسترس می‌باشند.

شبکه بیزین: شبکه بیزین یا شبکه‌های اعتقادی، جزئی از مدل‌های گرافیکی احتمالی محسوب می‌شوند (Lerner و همکاران، ۲۰۰۰). ساختار یک شبکه بیزین در واقع یک نمایش نموداری از اثرات متقابل متغیرهایی است که باید مدل‌بندی شوند، به طوری که علاوه بر کیفیت رابطه بین متغیرهای مسئله، کمیت ارتباط بین این متغیرها را نیز به نمایش می‌گذارد (Davies، ۲۰۰۷). این روش بر مبنای محاسبات

فرض می‌شود که رابطه میان متغیرهای وابسته و مستقل توسط یک تابع جبری نظیر $f(x)$ به علاوه یک مقداری خطا یا اختلال^۳ مشخص می‌شود. همانند رابطه زیر؛

$$f(x) = w^T \cdot \phi(x) + b \quad (۳)$$

$$y = f(x) + \text{noise} \quad (۴)$$

هدف اصلی، پیدا کردن فرم تابع $f(x)$ است. این کار با آموزش مدل SVM روی یک مجموعه‌ای از داده‌ها تحت عنوان مجموعه آموزش که شامل فرایندی به منظور بهینه‌سازی تابع خطا است، محقق می‌شود. در این مطالعه مدل‌های رگرسیونی SVM نوع دوم با نام SVM - ε به دلیل کاربرد گسترده آن در مسائل رگرسیونی مورد استفاده قرار گرفت. در این مدل، تابع خطا به صورت زیر تعریف می‌شود.

$$\frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (۵)$$

تابع خطای فوق بایستی با توجه به محدودیت‌های زیر کمینه شود (Hamel, ۲۰۱۱).

$$W^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (۶)$$

$$y_i - W^T \phi(x_i) - b \leq \varepsilon + \xi_i$$

$$\xi_i \cdot \xi_i^* \geq 0 \quad N = 1, 2, 3, \dots$$

که در آن‌ها، C ثابت گنجایش که به صورت عددی ثابت و مثبت بوده و به عنوان عامل جریمه در هنگام اتفاق افتادن خطای آموزشی مدل عمل می‌کند، بردار ضرایب، W^T ترانهاده بردار ضرایب، ξ_i^* و ξ_i متغیرهای کمبود برای مشخص کردن حدود بالا و پایین خطای آموزشی مرتبط با مقدار خطای مجاز ε ، b ثابت مشخصه‌های تابع رگرسیونی، N تعداد نمونه‌ها و ϕ تابع هسته^۴ است. در نهایت تابع SVM رگرسیونی را می‌توان به شکل زیر بازنویسی نمود.

$$f(x_i) = \sum_{i=1}^N \bar{\alpha}_i \phi(x_i)^T \cdot \phi(x) + b \quad (۷)$$

که در آن، $\bar{\alpha}_i$ میانگین ضرایب لاگرانژ می‌باشد. ماشین‌های بردار پشتیبان برای حل مسائل غیر خطی، ابعاد مسئله را از طریق توابع هسته تغییر می‌دهند.

الگوریتم‌های یادگیری ساختاری شبکه بیزین به دو دسته شامل الگوریتم‌های یادگیری بر مبنای محدودیت و الگوریتم‌های یادگیری بر مبنای امتیاز تقسیم می‌شوند. که دسته اول بر اساس استقلال شرطی و وابستگی بین متغیرها، به وسیله آزمون‌های آماری به دست می‌آید (مانند الگوریتم‌های PC^۱ و NPT^۲). در روش‌های یادگیری بر مبنای امتیاز نیز همه روابط ممکن بین گره‌ها را ارزیابی می‌کند و یک ساختار با بالاترین امتیاز را به عنوان ساختار مورد نظر انتخاب می‌کند (Sadeghi Hesar, ۲۰۱۲). دو الگوریتم PC و NPT با توجه به سادگی، بیشترین استفاده را در آموزش شبکه‌های بیزین برای آموزش ساختار مدل‌سازی دارند. در این پژوهش از الگوریتم PC برای آموزش در شبکه‌های بیزین، استفاده شد. ایده اصلی در این الگوریتم، استخراج مجموعه‌ای از استقلال و وابستگی عبارات شرطی به وسیله آزمون‌های آماری است که در آن استقلال شرطی بین هر جفت متغیر وجود ندارد و جهت کمان‌ها بر اساس استقلال شرطی تعیین می‌شود. شبکه بیزین علاوه بر پیش‌بینی متغیر هدف با استفاده از وضعیت متغیرهای ورودی، قادر است با در دست داشتن وضعیت متغیر پیش‌بینی شونده، میزان تاثیر هر یک از متغیرهای ورودی را بر خروجی‌های مدل تعیین کند (Aguilera و همکاران، ۲۰۱۱). بنابراین در پژوهش حاضر، این مدل انتخاب شد. برای اجرای شبکه بیزین از نرم‌افزار GeNIe 2 استفاده شد که به راحتی در اختیار سایر محققین قرار می‌گیرد.

ماشین بردار پشتیبان: ماشین بردار پشتیبان یک سیستم کارآمد یادگیری با نظارت است که برای دسته‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. این روش بر مبنای تئوری بهینه‌سازی مقید بنا شده است. به طوری که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و در نهایت منجر به یک جواب بهینه کلی می‌شود (Vapnik, ۱۹۹۸). در مدل رگرسیونی SVM تابعی مرتبط با متغیر وابسته y که به مجموعه‌ای از متغیرهای مستقل x بستگی دارد، تخمین زده می‌شود. همانند دیگر مسائل رگرسیونی

^۳ Noise
^۴ Kernel

^۱ Path Condition
^۲ Necessary Path Condition

ساتکلیف (NS^1) و اریب^۶ به صورت زیر استفاده شد. بهترین مقادیر برای این چهار معیار به ترتیب یک، صفر، یک و صفر می‌باشند.

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

$$R^2 = \frac{\left(\frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2}} \right)^2}{0 \leq R^2 \leq 1} \quad (9)$$

$$NS = 1 - \frac{\sum_{i=1}^N (X_i - Y_i)^2}{\sum_{i=1}^N (X_i - \bar{Y}_i)^2} \quad (10)$$

$$-\infty \leq NS \leq 1$$

$$Bias = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (11)$$

که در آن‌ها، X_i و Y_i به ترتیب مقادیر محاسباتی و مشاهداتی در گام زمانی N ام، تعداد گام‌های زمانی، \bar{X} و \bar{Y} نیز به ترتیب میانگین مقادیر مشاهداتی و محاسباتی می‌باشد.

با توجه به دامنه تغییرات زیاد داده‌ها وارد کردن آن‌ها، به صورت خام در مدل باعث کاهش سرعت و دقت شبکه می‌شود، بنابراین، لازم است به منظور افزایش کارایی مدل‌ها در ابتدا داده‌ها به شکل نرمال استاندارد تبدیل شوند (Xu و همکاران، ۲۰۰۷). بدین جهت برای نرمال‌سازی داده‌های ورودی از رابطه زیر استفاده شد.

$$X_n = 0.1 + 0.8 \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (12)$$

که در آن، X_n مقدار نرمال شده ورودی X_i و X_{\min} و X_{\max} به ترتیب حداکثر و حداقل داده‌ها می‌باشد.

نتایج و بحث

در این تحقیق، به منظور برآورد و تحلیل میزان مواد جامد محلول در رودخانه کشکان از داده‌های کیفیت مربوط به ایستگاه کشکان پلدختر شامل متغیرهای هیدروژن کربنات (HCO_3)، کلرید (Cl)،

انتخاب هسته برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی، بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع هسته‌ای را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. می‌توان توابع مختلف هسته برای ساخت انواع مختلف SVM رگرسیونی را به کار برد. با توجه به این که از پرکاربردترین توابع هسته، هسته خطی، پایه شعاعی و چندجمله‌ای می‌باشد (Basak و همکاران، ۲۰۰۷؛ Liu، ۲۰۱۱)، در این تحقیق نیز از این سه تابع هسته استفاده شد، که روابط ریاضی آن‌ها در جدول ۱ ارائه شده است.

جدول ۱ - توابع هسته مختلف

رابطه ریاضی	نوع کرنل
$K(x, x') = (x \cdot x')$	هسته خطی ^۱
$K(x, x') = ((x \cdot x') + 1)^d$	هسته چندجمله‌ای ^۲
$K(x, x') = \exp\left(-\frac{\ x - x'\ ^2}{2\sigma^2}\right)$	هسته توابع پایه شعاعی (RBF) ^۳

به جای تابع پیچیده $\phi(x)$ و انجام محاسبات در فضایی با ابعاد بالا از تابع ساده‌تری به صورت جدول ۱ استفاده می‌شود. ساده‌ترین نوع تابع هسته، هسته خطی می‌باشد. هسته چند جمله‌ای یک هسته مورد توجه برای مدل‌های غیر خطی می‌باشد و به همراه هسته پایه شعاعی اکنون اهمیت زیادی پیدا کرده‌اند و بیشتر به صورت گوسی^۴ (جدول ۱) به کار می‌روند. در روابط بالا d و σ به ترتیب، درجه هسته چند جمله‌ای و عرض هسته تابع پایه شعاعی است. برای اجرای مدل ماشین بردار پشتیبان در محیط MATLAB 2013 کد نویسی انجام شده و کارایی هسته‌های ذکر شده مورد بررسی و ارزیابی قرار گرفت.

معیارهای ارزیابی: به منظور ارزیابی دقت و کارایی مدل‌ها، از معیارهای ضریب تبیین (R^2)، ریشه میانگین مربعات خطا ($RMSE^5$)، ضریب نش-

¹ Liner

² Polynomial

³ Radial Basis Function

⁴ Gaussian

⁵ Root Mean Square Error

⁶ Nash Sutcliffe

⁷ Bias

مدل‌های شبکه بیزین، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان استفاده شد. جدول ۲ مشخصات آماری داده‌های کیفیت آب رودخانه کشکان در ایستگاه پلدختر را نشان می‌دهد.

سولفات (SO_4)، منیزیم (Mg)، کلسیم (Ca)، سدیم (Na)، هدایت الکتریکی (EC)، pH و دبی جریان در مقیاس زمانی ماهانه طی دوره آماری (۱۳۹۵-۱۳۷۰) به‌عنوان ورودی و متغیر TDS به‌عنوان خروجی

جدول ۲- مشخصات آماری متغیرهای کیفی رودخانه کشکان

پارامتر	واحد	تعداد رکورد	حداقل	حداکثر	میانگین	انحراف معیار	ضریب تغییرات	چولگی
TDS	میلی‌گرم بر لیتر	۳۰۰	۲۰۵	۱۴۳۲	۴۴۵/۷	۱۶۵/۷	۰/۳۷	۲/۶۴
EC	میلی‌گرم بر لیتر	۳۰۰	۳۳۰	۲۱۷۰	۶۸۲/۸۵	۲۴۶/۴۴	۰/۳۶	۲/۶۷
Hco3	میلی‌گرم بر لیتر	۳۰۰	۱/۵۵	۵	۳/۱	۰/۷۷	۰/۲۵	۰/۱۷
Cl	میلی‌گرم بر لیتر	۳۰۰	۰/۸	۱۴/۱	۲/۷۵	۲/۶	۰/۹۴	۳/۱
So4	میلی‌گرم بر لیتر	۳۰۰	۰	۹/۱۹	۱/۸۹	۱/۷۹	۰/۹۵	۱/۸
Ca	میلی‌گرم بر لیتر	۳۰۰	۲	۵	۳/۲	۰/۸۸	۰/۲۸	۰/۲۲
Mg	میلی‌گرم بر لیتر	۳۰۰	۱	۵/۴	۲/۲۷	۱/۱۳	۰/۵	۱/۳
Na	میلی‌گرم بر لیتر	۳۰۰	۱	۲/۱	۱/۵	۰/۵۶	۰/۳۶	۰/۴۱
pH	-	۳۰۰	۷	۸	۷/۸۳	۰/۳۶	۰/۰۵	-۲/۱۶
Q	میلی‌گرم بر ثانیه	۳۰۰	۰	۲۹۹/۸	۴۱/۶۱	۵۰/۲	۱/۲	۲/۸

گام‌به‌گام به‌ترتیب اولویت تأثیر بر مقدار جامدات محلول رودخانه (ضریب همبستگی بین پارامترهای ورودی و مقدار کل مواد جامد محلول) به‌عنوان ورودی مدل‌ها تعیین شدند. نتایج به‌دست آمده در جدول ۳ ارائه شده است. بر اساس جدول همبستگی، هدایت الکتریکی و pH به‌ترتیب بیشترین و کمترین همبستگی را با کل مواد جامد محلول در آب دارند. برای مدل‌سازی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد داده‌ها به‌صورت تصادفی برای صحت‌سنجی انتخاب شدند. در جدول ۴ الگوهای ورودی مدل مورد استفاده در مدل‌ها نشان داده شده است.

یکی از مراحل مهم در مدل‌سازی، انتخاب ترکیبی مناسب از متغیرهای ورودی به مدل می‌باشد. به‌عبارت دیگر انتخاب و ترکیب ورودی‌های مناسب و تاثیرگذار در پدیده برای آموزش ماهیت ساز و کار حاکم بر پدیده باعث توسعه و بهبود عملکرد مدل خواهد شد. بدین جهت برای تعیین پارامترهای موثر و الگوهای ورودی مدل‌ها برای برآورد کل مواد جامد محلول در آب (TDS) ابتدا ماتریس همبستگی بین کل املاح جامد محلول در آب و متغیرهای ورودی ایستگاه مورد مطالعه تشکیل و سپس پارامترهایی که دارای همبستگی معنی‌دار بودند، بر اساس رگرسیون

جدول ۳- ماتریس همبستگی متغیرهای کیفی رودخانه کشکان

Q	pH	Na	Mg	Ca	So4	Cl	Hco3	EC	TDS	
									۱	TDS
								۱	۰/۹۸۸	EC
							۱	۰/۲۲۴	۰/۲۱۳	Hco3
						۱	۰/۹۱۶	۰/۷۸۰	۰/۷۶۹	Cl
					۱	۰/۱۴۳	-۰/۲۸۴	۰/۵۵	۰/۵۵۳	So4
				۱	۰/۶۴۴	۰/۴۵۵	۰/۲۳۴	۰/۷۴۹	۰/۷۳۷	Ca
			۱	۰/۱۸۹	۰/۱۱۵	۰/۶۴۰	۰/۳۳۹	۰/۵۷۶	۰/۵۸۸	Mg
		۱	۰/۰۴۸	-۰/۰۰۹	۰/۳۱۰	۰/۴۳۰	-۰/۲۵۷	۰/۳۸۱	۰/۳۶۲	Na
	۱	۰/۴۹۵	-۰/۲۱۱	-۰/۱۲۶	۰/۱۹۴	-۰/۰۲۰	-۰/۳۵۴	۰/۰۴۵۸	۰/۰۴۸	pH
۱	۰/۰۷۷	۰/۰۰۸	-۰/۰۹۷	-۰/۰۲۲	-۰/۰۸۴	-۰/۱۱۷	-۰/۱۷۳	-۰/۲۰۵	-۰/۱۹۹	Q

جدول ۴- الگوهای ورودی مدل‌ها

ردیف	الگوی ورودی مدل
۱	TDS=f{EC}
۲	TDS=f{EC, Cl}
۳	TDS=f{EC, Cl, Ca}
۴	TDS=f{EC, Cl, Ca, Mg}
۵	TDS=f{EC, Cl, Ca, Mg, SO4}
۶	TDS=f{EC, Cl, Ca, Mg, SO4, Na}
۷	TDS=f{EC, Cl, Ca, Mg, SO4, Na, Hco3}
۸	TDS=f{EC, Cl, Ca, Mg, SO4, Na, Hco3, Q}
۹	TDS=f{EC, Cl, Ca, Mg, SO4, Na, Hco3, Q, pH}

کمترین ریشه میانگین مربعات خطا (میلی‌گرم بر لیتر) $RMSE=0/092$ ، بیشترین ضریب نش-ساتکلیف $NS=0/707$ و کمترین اریب $Bias=0/008$ در مرحله صحت‌سنجی عملکرد بهتری نسبت به سایر ساختارها داشته و توانسته است میزان کل مواد جامد محلول در آب منطقه مورد مطالعه را با دقت مناسبی شبیه‌سازی نماید که در نتایج Ghorbani و همکاران (۲۰۱۶) نیز به آن اشاره شده است.

در شکل ۳- الف تغییرات مقادیر مشاهداتی و محاسباتی نسبت به زمان نشان داده شده است. شکل ۳- ب نمودار میزان تطابق مقادیر تخمینی با مقادیر محاسباتی را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، بیشتر مقادیر در اطراف خط نیم‌ساز پراکنده شده‌اند که حاکی از نزدیک بودن مقادیر برآوردی به مقادیر مشاهداتی می‌باشد. ولی در خصوص مقادیر بیشینه، مدل عملکرد مناسبی نداشته و می‌توان اظهار نمود که مدل شبکه بیزین در تخمین مقادیر بیشینه، ضعیف عمل کرده و مقادیر را کمتر از مقدار واقعی تخمین زده است که در شکل مشهود است؛ در بیان آن می‌توان اظهار نمود، شبکه بیزین ارتباط بین هر یک از متغیرها به صورت تصادفی بر اساس استقلال شرطی و وابستگی بین متغیرها و بر پایه احتمالات، مشخص می‌کند و شبکه به‌خوبی تعمیم داده نشده است. این موضوع با نتایج Dehghani و همکاران (۲۰۱۶) هم‌خوانی دارد.

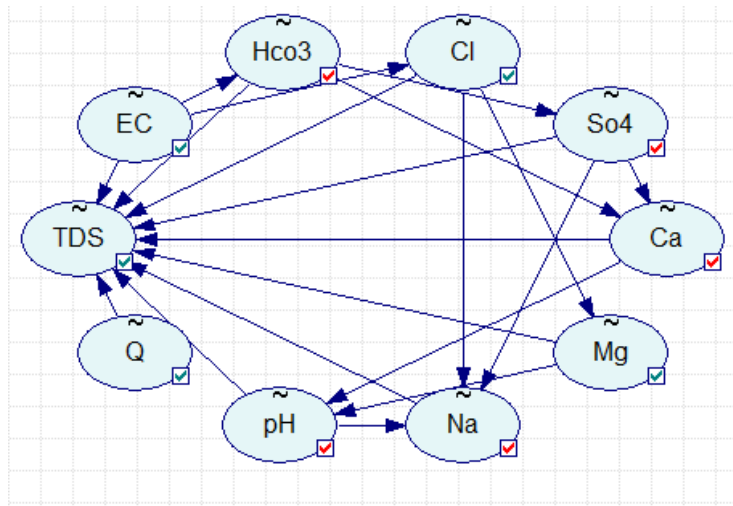
در این پژوهش، پس از نرمال‌سازی داده‌های ورودی، تمام ترکیب‌های الگوهای ورودی برای هر یک از روش‌های مورد استفاده قرار گرفتند. لازم به ذکر است که این مراحل در مدل SVM به‌ازای هر سه هسته (۲۷ الگو) و برای برنامه‌ریزی بیان ژن به‌ازای هر دو عملگر (۱۸ الگو) و برای شبکه بیزین به‌ازای هر نه الگوی ورودی (جدول ۳) انجام گرفت. در پایان، برای مقایسه عملکرد مدل‌ها در دو بخش واسنجی و صحت‌سنجی از معیارهای ارزیابی استفاده شد.

نتایج شبکه بیزین: در این پژوهش، برای آموزش ساختاری شبکه از الگوریتم یادگیری جستجوی ممنوعه (PC^1) و نیز برای آموزش پارامتری شبکه، با تنظیم دو متغیر سطح معنی‌داری^۲ و حداکثر اندازه مجاورت^۳ با توجه به تاثیر پارامترها بر یکدیگر استفاده شد. شبکه بیزین به‌ازای هر نه الگوی ورودی ارائه شده در جدول ۳ اجرا شد. در شکل ۲ بهترین ساختار شبکه بیزین نشان داده شده است. هدف اصلی از این روش، یافتن ارتباط بین میزان کل مواد جامد محلول و پارامترهای تأثیرگذار بر آن می‌باشد که در نهایت منجر به ارائه رابطه ۱۳ شد. همچنین، در جدول ۵ نتایج حاصل از شبکه بیزین نشان داده شده است. همان‌طور که مشاهده می‌شود، ساختارهای ترکیبی عملکرد بهتری را ارائه کردند به طوری که ساختار ترکیبی شماره ۹ با بیشترین دقت $R^2=0/8031$ ،

¹ Path Condition

² Significance Level

³ Max Adjacency Size

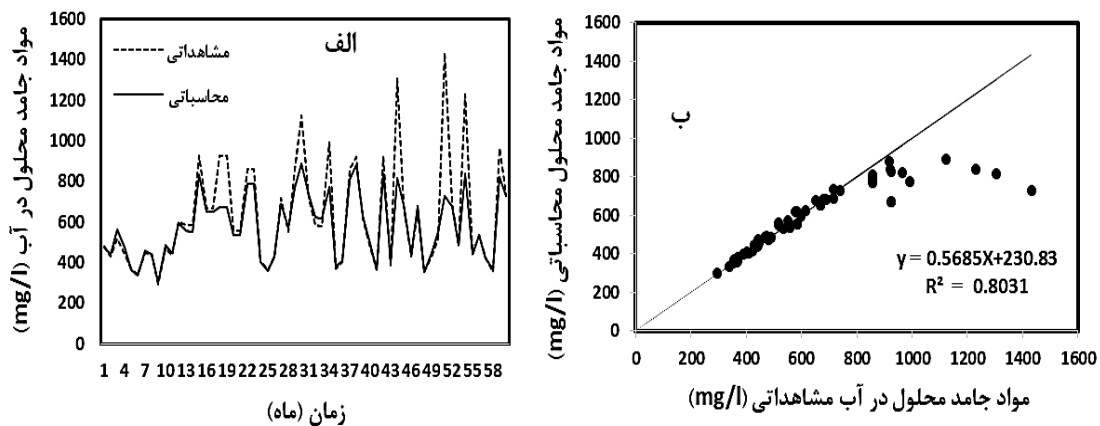


شکل ۲- ساختار شبکه بیزین به کار رفته برای شبیه‌سازی کل مواد جامد محلول در آب

جدول ۵ - نتایج مدل شبکه بیزین

صحت سنجی				آموزش				الگوها	ردیف
RMSE	NS	Bias	R ²	RMSE	NS	Bias	R ²		
۰/۰۹۹	۰/۶۸۱	۰/۰۰۹۸	۰/۷۷۳	۰/۱۲۴۲	۰/۶۵۲	۰/۰۱۵۴	۰/۷۴۵	M1	۱
۰/۰۹۸	۰/۶۸۴	۰/۰۰۹۶	۰/۷۷۵	۰/۱۲۴۰	۰/۷۵۹	۰/۰۱۵۴	۰/۷۴۸	M2	۲
۰/۰۹۸	۰/۶۸۹	۰/۰۰۹۶	۰/۷۷۸	۰/۱۲۳۹	۰/۷۶۵	۰/۰۱۵۴	۰/۷۵۲	M3	۳
۰/۰۹۷	۰/۶۹۳	۰/۰۰۹۴	۰/۷۸۱	۰/۱۲۳۷	۰/۷۷۱	۰/۰۱۵۳	۰/۷۶۷	M4	۴
۰/۰۹۵	۰/۶۹۶	۰/۰۰۹۰	۰/۷۸۶	۰/۱۲۳۶	۰/۷۷۶	۰/۰۱۵۳	۰/۷۷۲	M5	۵
۰/۰۹۵	۰/۶۹۹	۰/۰۰۹۰	۰/۷۹۲	۰/۱۲۳۴	۰/۷۸۰	۰/۰۱۵۲	۰/۷۸۵	M6	۶
۰/۰۹۳	۰/۷۰۲	۰/۰۰۸۶	۰/۷۹۷	۰/۱۲۳۳	۰/۷۸۴	۰/۰۱۵۲	۰/۷۸۹	M7	۷
۰/۰۹۲	۰/۷۰۴	۰/۰۰۸	۰/۸۰۱	۰/۱۲۳۰	۰/۷۸۹	۰/۰۱۵۱	۰/۷۹۳	M8	۸
۰/۰۹۲	۰/۷۰۷	۰/۰۰۸	۰/۸۰۳	۰/۱۲۲۹	۰/۷۹۲	۰/۰۱۵۱	۰/۷۹۶	M9	۹

$$TDS = 0.973 \times EC + 0.015 \times Hco3 + 0.071 \times Cl + 0.024 \times So4 - 0.014 \times Ca - 0.003 \times Mg - 0.016 \times Na + 0.013 \times pH + 0.001 \times Q + Normal(0.0218, 0.018) \quad (13)$$



شکل ۳- نمودار حاصل از مقادیر بهینه مدل شبکه بیزین برای داده‌های ثبت‌شده مرحله صحت‌سنجی، الف- مقادیر محاسباتی و مشاهداتی نسبت به زمان و ب- نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

پژوهش، مدل برنامه‌ریزی بیان ژن یک بار بر اساس عملگرهای ریاضی پیش‌فرض مدل GeneXpro (عملگرهای F2) و بار دیگر با استفاده از چهار عملگر اصلی (F1)، روابط ۱۳ و ۱۴ مورد ارزیابی قرار گرفت تا کارایی مدل به ازای عملگرهای ساده و پیچیده مورد بررسی و در نهایت عملگر مناسب انتخاب شود. نحوه و دلیل انتخاب این نوع عملگرها با توجه به مطالعات Ghorbani و همکاران (۲۰۱۲) صورت گرفت.

نتایج برنامه‌ریزی بیان ژن: در این پژوهش، با توجه به قابلیت و توانایی برنامه‌ریزی بیان ژن در انتخاب متغیرهای موثر و حذف متغیرهایی با تاثیر کمتر و ارائه رابطه صریح استفاده از آن برای تخمین میزان مواد جامد محلول در آب رودخانه کشکان پلدختر مد نظر قرار گرفت. یکی از گام‌های مهم در فرایند اجرای مدل برنامه‌ریزی بیان ژن انتخاب عملگرهای اصلی، به‌منظور ساخت درخت تجزیه می‌باشد. در این

$$F1 = \{+, -, *, /\} \tag{13}$$

$$F2 = \{+, -, *, /, \sqrt{\quad}, \exp, \ln, 2, 3, \sqrt[3]{\quad}, \sin, \cos, \text{atan}\} \tag{14}$$

تعداد ژن‌ها و کروموزوم‌ها، به‌منظور اجرای مدل برنامه‌ریزی بیان ژن، بایستی تعیین شوند. با توجه به مطالعات Ferreira، پاسخ بهینه با انتخاب دو یا سه ژن حاصل می‌شود. همچنین، وی در مطالعاتش نشان داد که بهترین جمعیت، از ۳۰ تا ۵۰ کروموزوم تشکیل می‌شود (Ferreira, ۲۰۰۱). بنابراین، در پژوهش حاضر، از سه ژن و ۳۰ کروموزوم استفاده شد. نتایج حاصل مدل‌سازی مقدار مواد جامد محلول با استفاده از برنامه‌ریزی بیان ژن در جدول ۶ ارائه شده است.

همان‌طور که در جدول مشاهده می‌شود، ترکیب نه با عملگر F1 در هر دو مرحله آموزش و صحت‌سنجی دارای بهترین عملکرد با ضریب تبیین $R^2 = 0.970$ ، ریشه میانگین مربعات خطا

RMSE=۰/۰۳۳ (میلی‌گرم بر لیتر) ضریب نش-ساتکلیف NS=۰/۹۶۰ و اریب Bias=۰/۰۰۱۱ می‌باشد. نتایج به‌دست آمده نشان داد که برنامه‌ریزی بیان ژن با عملگر F1 شامل چهار عملگر اصلی نسبت به دیگر عملگرها از دقت بالاتری برخوردار بوده است. به‌عبارت دیگر اضافه کردن عملگرهای پیچیده نه تنها باعث افزایش دقت مدل نشده، بلکه در مواردی سبب کاهش دقت مدل به‌دست آمده شده است، که با نتایج Dehghani و Ghorbani (۲۰۱۷) هم‌خوانی دارد.

رابطه استخراج شده از مدل برنامه‌ریزی بیان ژن با عملگر F1 برای ایستگاه کشکان-پلدختر، در رابطه (۱۵) ارائه شده است.

جدول ۶- روش برنامه‌ریزی بیان ژن با استفاده از سه مجموعه عملگر ریاضی

صحت‌سنجی				آموزش				عملگر	الگوها	ردیف
RMSE	NS	Bias	R ²	RMSE	NS	Bias	R ²			
-۰/۳۵۷	۰/۹۲۷	-۰/۰۱۳	۰/۹۳۴	-۰/۱۱۴	۰/۸۷۶	-۰/۰۱۳	۰/۹۱۷	F1	M1	۱
-۰/۳۵۷	۰/۹۲۶	-۰/۰۱۳	۰/۹۲۷	-۰/۱۰۹۵	۰/۸۷۶	-۰/۰۱۲	۰/۹۱۳	F2		
-۰/۳۵۵	۰/۹۳۰	-۰/۰۱۳	۰/۹۴۹	-۰/۱۰۱۵	۰/۸۷۹	-۰/۰۰۳	۰/۹۱۹	F1	M2	۲
-۰/۳۵۵	۰/۹۲۹	-۰/۰۱۳	۰/۹۴۷	-۰/۱۰۴۴	۰/۸۷۶	-۰/۰۰۹	۰/۹۱۸	F2		
-۰/۳۵۳	۰/۹۴۳	-۰/۰۰۱۲	۰/۹۵۶	-۰/۰۹۸۵	۰/۸۸۲	-۰/۰۰۹۷	۰/۹۲۰	F1	M3	۳
-۰/۳۵۵	۰/۹۴۰	-۰/۰۰۱۳	۰/۹۴۷	-۰/۰۹۸۴	۰/۸۸۱	-۰/۰۰۹۷	۰/۹۲۰	F2		
-۰/۳۵۲	۰/۹۴۷	-۰/۰۰۱۲	۰/۹۵۸	-۰/۰۹۷	۰/۸۸۹	-۰/۰۰۹۴	۰/۹۲۱	F1	M4	۴
-۰/۳۵۰	۰/۹۴۳	-۰/۰۰۱۲	۰/۹۵۷	-۰/۰۹۷	۰/۸۸۵	-۰/۰۰۹۵	۰/۹۲۰	F2		
-۰/۳۴۷	۰/۹۴۷	-۰/۰۰۱۲	۰/۹۶۰	-۰/۰۹۶	۰/۸۹۰	-۰/۰۰۹۳	۰/۹۳۱	F1	M5	۵
-۰/۳۴۷	۰/۹۴۹	-۰/۰۰۱۲	۰/۹۵۷	-۰/۰۹۷	۰/۸۹۰	-۰/۰۰۹۴	۰/۹۲۹	F2		
-۰/۳۴۵	۰/۹۵۳	-۰/۰۰۱۲	۰/۹۶۶	-۰/۰۹۴	۰/۸۹۲	-۰/۰۰۸۹	۰/۹۳۳	F1	M6	۶
-۰/۳۴۵	۰/۹۵۱	-۰/۰۰۱۲	۰/۹۶۱	-۰/۰۹۷	۰/۸۹۰	-۰/۰۰۹۴	۰/۹۳۳	F2		
-۰/۳۴۲	۰/۹۵۶	-۰/۰۰۱۲	۰/۹۶۵	-۰/۰۹۴	۰/۸۹۵	-۰/۰۰۸۸	۰/۹۴۲	F1	M7	۷
-۰/۳۴۲	۰/۹۵۴	-۰/۰۰۱۲	۰/۹۶۵	-۰/۰۹۴	۰/۸۸۹	-۰/۰۰۸۹	۰/۹۴۳	F2		
-۰/۳۳۷	۰/۹۵۶	-۰/۰۰۱۱	۰/۹۶۷	-۰/۰۹۳	۰/۸۹۹	-۰/۰۰۸۷	۰/۹۴۸	F1	M8	۸
-۰/۳۳۶	۰/۹۵۶	-۰/۰۰۱۱	۰/۹۶۵	-۰/۰۹۳	۰/۸۹۸	-۰/۰۰۸۷	۰/۹۴۷	F2		
-۰/۳۳۴	۰/۹۵۹۸	-۰/۰۰۱۱	۰/۹۷۰۳	-۰/۰۹۱	۰/۹۳۳	-۰/۰۰۸۳	۰/۹۴۸	F1	M9	۹
-۰/۳۳۵	۰/۹۵۷	-۰/۰۰۱۱	۰/۹۶۷	-۰/۰۹۱	۰/۹۳۳	-۰/۰۰۸۳	۰/۹۴۸	F2		

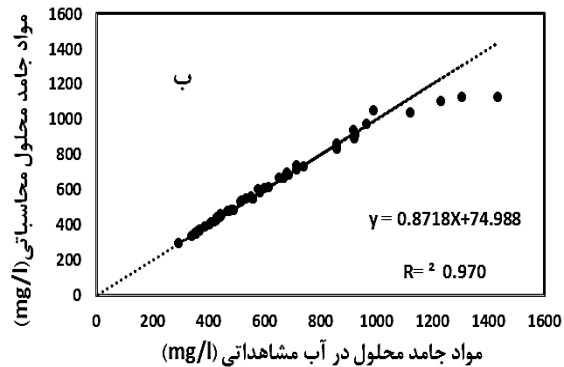
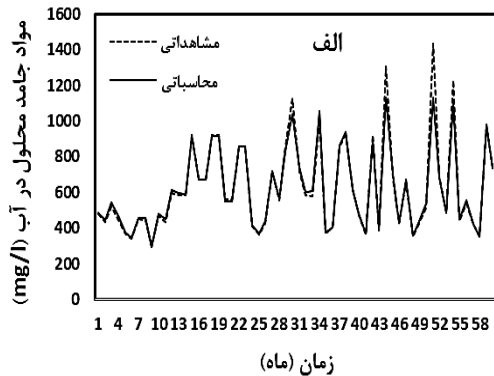
$$TDS = \left(\left(\left(\left(\left(5/115 - [Ca] \right) - ([Na]/[EC]) \right) / \left(\left(2/798 + [Na] \right) / [Mg] \right) \right) / \left(9/002 / ([So4] \times [EC]) \right) \right) + \left(\left([Ca] / \left(\left([Ca] - 6/082 \right) \times 3/425 \right) + ([Hco3] + [So4]) \right) \right) \times ([Mg] \times 0/477) \right) + [EC] \quad (15)$$

شده است که در تبیین آن می‌توان نمود شبکه در عملگر F1 به‌خوبی تعمیم داده شده است. ولی در برآورد برخی از مقادیر بیشینه از دقت آن کاسته شده است. در بیان این نتایج می‌توان اظهار نمود معادله حاصل از برنامه‌ریزی بیان ژن، از ترکیب تصادفی مجموع ترمینال‌ها و توابع به‌دست می‌آید. لذا، از مجموعه عملگرهای انتخابی از جمله \ln ، \cos ، \sin و ... برای استخراج رابطه‌ای استفاده می‌شود که ممکن است رابطه بین ورودی‌ها و خروجی‌های آن خطی باشد که این امر باعث کاهش دقت مدل می‌شود. در

نمودار مقادیر بهینه محاسباتی و مشاهداتی برنامه‌ریزی بیان ژن در شکل ۴ نشان داده شده است. بر اساس شکل ۴-ب، مشاهده می‌شود که عملکرد مدل مناسب بوده به‌طوری که در بیشتر موارد مقادیر تخمینی و مشاهداتی روی خط نیمساز ($x=y$) قرار گرفته که نشان‌دهنده برابر بودن مقادیر مشاهداتی به مقادیر محاسباتی می‌باشد. همان‌طور که در شکل‌ها مشاهده می‌شود، این مدل در برآورد مقادیر کمینه عملکرد مناسبی نشان داده است. به‌طوری که مقادیر برآوردی نزدیک به مقادیر واقعی آن‌ها تخمین زده

جهت برآورد مواد جامد محلول در آب رودخانه کشکان پیشنهاد شد.

این پژوهش، عملگرهای $\sqrt{\cdot}$, \ln , \cos , \sin , Exp و Atan کارایی مناسب جهت افزایش دقت مدل نداشته و نیز چهار عملگر اصلی ریاضی با توجه به دقت و سادگی،



شکل ۴- نمودار حاصل از مقادیر بهینه مدل برنامه‌ریزی بیان ژن برای داده‌های ثبت‌شده مرحله صحت‌سنجی، الف- مقادیر محاسباتی و مشاهده‌ای نسبت به زمان و ب- نمودار پراکنش میان مقادیر مشاهده‌ای و محاسباتی

شکل ۵، نمودار بهترین مدل حاصل‌شده برای داده‌های بخش صحت‌سنجی نشان داده شده است.

با توجه به شکل ۵، مشاهده می‌شود که مدل ماشین بردار پشتیبان از عملکرد مناسبی برخوردار بوده، به طوری که مدل در تخمین و پیش‌بینی اکثر مقادیر دقت قابل قبولی داشته و مقادیر پیش‌بینی شده را نزدیک به مقدار واقعی‌شان تخمین زده است. ولی در تخمین مقادیر بیشینه در بعضی از نقاط عملکرد دقیقی از خود نشان نداده است که این نتایج با نتایج Eskandari و همکاران (۲۰۱۲) و Ggorbani و Dehghani (۲۰۱۷) هم‌خوانی دارد.

در تبیین نتایج بالا می‌توان بیان نمود که ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده می‌نماید. بنابراین با به‌کارگیری روش یادگیری با نظارت با استفاده از توابع پایه شعاعی در پیش‌بینی، باعث می‌شود تخمین پارامتر سرعت و دقت بالاتر و خطای کمتری نسبت به سایر هسته‌ها داشته باشد که یکی از خصوصیات بارز توابع پایه شعاعی به شمار می‌رود.

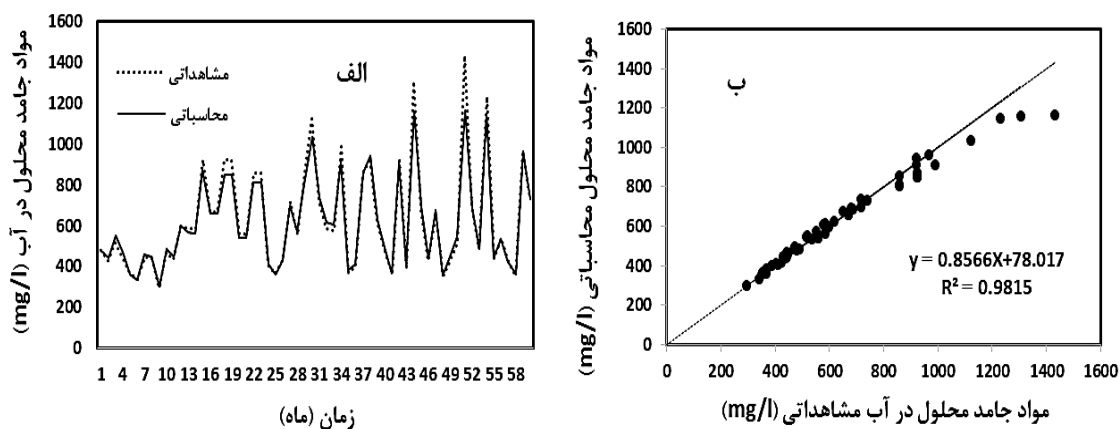
نتایج ماشین بردار پشتیبان: در اجرای مدل ماشین بردار پشتیبان، عملکرد مدل به ازای توابع مختلف هسته شامل هسته چندجمله‌ای، خطی و توابع پایه شعاعی و با استفاده از ساختارهای مختلف ورودی مورد بررسی و ارزیابی قرار گرفت. نتایج حاصل از ساختارها به ازای هسته‌های مذکور که در مجموع ۲۷ الگو می‌باشند، در جدول ۷ آورده شده است. با توجه به جدول مشاهده می‌شود، ساختار مدل ترکیبی شماره ۹ با تابع هسته پایه شعاعی (RBF) با بیشترین ضریب تبیین $R^2 = 0.982$ ، کمترین ریشه میانگین مربعات خطا $RMSE = 0.32 \text{ (mg/lit}^{-1}\text{)}$ ، بیشترین ضریب نش-ساتکلیف $NS = 0.963$ و اریب $0.01 =$ Bias در مرحله صحت‌سنجی جواب بهینه‌تری نسبت به سایر ساختارها داشته است.

قابل ذکر است که در برخی از الگوها نتایج هسته چندجمله‌ای نزدیک به هسته پایه شعاعی مشاهده شد؛ اما در کل نتایج ماشین بردار با تابع هسته پایه شعاعی عملکرد و دقت بیشتری داشت که با نتایج Haghiahi و همکاران (۲۰۱۸) هم‌خوانی دارد. در

جدول ۷- روش ماشین بردار پشتیبان با استفاده از سه تابع کرنل پایه شعاعی، کرنل چندجمله‌ای و کرنل خطی

ردیف	الگوها	عملگر	آموزش				صحت سنجی			
			RMSE	NS	Bias	R ²	RMSE	NS	Bias	R ²
۱	M1	RBF	۰/۹۳۶	۰/۸۹۳	۰/۰۸۷	۰/۹۶۲	۰/۰۳۴۳	۰/۹۴۰	۰/۰۱۲	۰/۹۶۲
		Poly1	۰/۹۳۵	۰/۸۹۳	۰/۰۸۹	۰/۹۵۸	۰/۰۳۵۲	۰/۹۳۸	۰/۰۱۲	۰/۹۵۸
		Lin ²	۰/۹۳۳	۰/۸۸۹	۰/۰۹۲	۰/۹۳۹	۰/۰۳۵۱	۰/۹۳۳	۰/۰۱۲	۰/۹۳۹
۲	M2	RBF	۰/۹۴۵	۰/۸۹۳	۰/۰۸۴	۰/۹۶۳	۰/۰۳۴۳	۰/۹۴۱	۰/۰۱۲	۰/۹۶۳
		Poly	۰/۹۴۴	۰/۸۹۳	۰/۰۸۷	۰/۹۶۳	۰/۰۳۴۳	۰/۹۳۵	۰/۰۱۲	۰/۹۶۳
		Lin	۰/۹۴۳	۰/۸۸۹	۰/۰۸۹	۰/۹۵۶	۰/۰۳۵	۰/۹۳۰	۰/۰۱۲	۰/۹۵۶
۳	M3	RBF	۰/۹۴۶	۰/۸۹۸	۰/۰۸۳	۰/۹۶۷	۰/۰۳۴۲	۰/۹۴۲	۰/۰۱۲	۰/۹۶۷
		Poly	۰/۹۴۳	۰/۸۹۸	۰/۰۸۵	۰/۹۶۵	۰/۰۳۴۲	۰/۹۳۸	۰/۰۱۲	۰/۹۶۵
		Lin	۰/۹۴۳	۰/۸۹۵	۰/۰۸۷	۰/۹۵۹	۰/۰۳۴۳	۰/۹۲۲	۰/۰۱۲	۰/۹۵۹
۴	M4	RBF	۰/۹۴۷	۰/۹۲۲	۰/۰۸۱	۰/۹۷۱	۰/۰۳۲۹	۰/۹۴۶	۰/۰۱۱	۰/۹۷۱
		Poly	۰/۹۴۶	۰/۹۲۰	۰/۰۸۵	۰/۹۶۶	۰/۰۳۳۵	۰/۹۴۶	۰/۰۱۱	۰/۹۶۶
		Lin	۰/۹۴۳	۰/۸۹۹	۰/۰۸۵	۰/۹۶۲	۰/۰۳۳۵	۰/۹۴۲	۰/۰۱۱	۰/۹۶۲
۵	M5	RBF	۰/۹۴۸	۰/۹۲۳	۰/۰۰۸	۰/۹۷۴	۰/۰۳۲۷	۰/۹۴۹	۰/۰۱۱	۰/۹۷۴
		Poly	۰/۹۴۷	۰/۹۲۰	۰/۰۰۸	۰/۹۷۱	۰/۰۳۲۷	۰/۹۴۷	۰/۰۱۱	۰/۹۷۱
		Lin	۰/۹۴۷	۰/۹۰۳	۰/۰۰۸	۰/۹۶۲	۰/۰۳۲۹	۰/۹۴۵	۰/۰۱۱	۰/۹۶۲
۶	M6	RBF	۰/۹۵۱	۰/۹۳۴	۰/۰۰۷۸	۰/۹۷۶	۰/۰۳۲۵	۰/۹۵۰	۰/۰۱۱	۰/۹۷۶
		Poly	۰/۹۴۷	۰/۹۳۳	۰/۰۰۸۱	۰/۹۷۱	۰/۰۳۲۷	۰/۹۴۹	۰/۰۱۱	۰/۹۷۱
		Lin	۰/۹۴۲	۰/۹۲۱	۰/۰۰۸۵	۰/۹۶۸	۰/۰۳۲۷	۰/۹۴۶	۰/۰۱۱	۰/۹۶۸
۷	M7	RBF	۰/۹۵۱	۰/۹۴۵	۰/۰۰۷۴	۰/۹۷۷	۰/۰۳۲۳	۰/۹۵۵	۰/۰۱۱	۰/۹۷۷
		Poly	۰/۹۴۹	۰/۹۴۱	۰/۰۰۷۸	۰/۹۶۹	۰/۰۳۲۳	۰/۹۵۳	۰/۰۱۰	۰/۹۶۹
		Lin	۰/۹۴۷	۰/۹۳۳	۰/۰۰۷۸	۰/۹۶۵	۰/۰۳۲۳	۰/۹۵۲	۰/۰۱۰	۰/۹۶۵
۸	M8	RBF	۰/۹۵۴	۰/۹۴۷	۰/۰۰۷۳	۰/۹۷۷	۰/۰۳۲۱	۰/۹۵۸	۰/۰۱۰	۰/۹۷۷
		Poly	۰/۹۵۱	۰/۹۴۵	۰/۰۰۷۴	۰/۹۷۳	۰/۰۳۲۱	۰/۹۵۷	۰/۰۱۰	۰/۹۷۳
		Lin	۰/۹۴۸	۰/۹۴۰	۰/۰۰۷۸	۰/۹۶۹	۰/۰۳۲۴	۰/۹۵۴	۰/۰۱۰	۰/۹۶۹
۹	M9	RBF	۰/۹۶۷	۰/۹۵۰۵	۰/۰۰۰۷	۰/۹۸۱۵	۰/۰۳۲۰	۰/۹۶۳	۰/۰۱۰	۰/۹۸۱۵
		Poly	۰/۹۶۵	۰/۹۴۹	۰/۰۰۰۷	۰/۹۸۰	۰/۰۳۲۲	۰/۹۶۰	۰/۰۱۰	۰/۹۸۰
		Lin	۰/۹۵۶	۰/۹۴۷	۰/۰۰۰۷۴	۰/۹۷۷	۰/۰۳۲۷	۰/۹۶۰	۰/۰۱۱	۰/۹۷۷

^۱ همان هسته Polynomial می‌باشد که در جدول به این صورت نوشته شده است. ^۲ همان هسته Liner می‌باشد که در جدول به این صورت نوشته شده است.



شکل ۵- نمودار حاصل از مقادیر بهینه مدل ماشین بردار پشتیبان برای داده‌های ثبت شده مرحله صحت‌سنجی، الف- مقادیر محاسباتی و مشاهداتی نسبت به زمان و ب- نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

ارزیابی و مقایسه عملکرد مدل‌ها: جواب بهینه هر یک از مدل‌های مورد استفاده در این پژوهش، در جدول ۸ جهت مقایسه آورده شده است. همان‌طور که در جدول ۸ مشاهده می‌شود، هر سه مدل بر اساس معیارهای ارزیابی توانسته‌اند با دقت قابل قبولی میزان کل مواد جامد در آب رودخانه کشکان پلدختر را

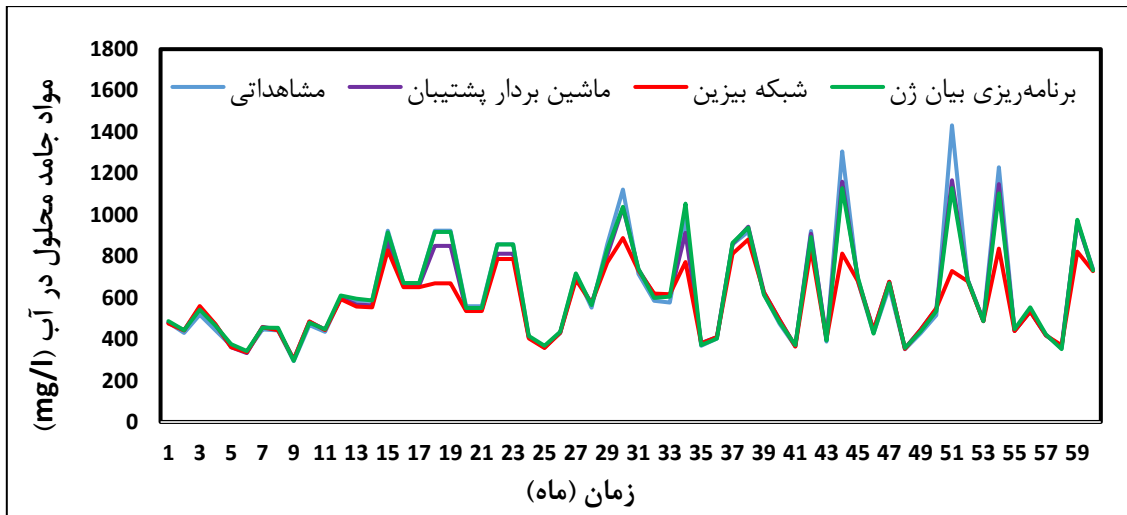
شبیه‌سازی کنند. همچنین، از بین مدل‌های به‌کار رفته مدل ماشین بردار پشتیبان با هسته پایه شعاعی از بیشترین دقت $R^2=0/9815$ و کمترین ریشه میانگین مربعات خطا (RMSE) $0/0320$ و کمترین مقدار اریب $Bias=0/001$ و بیشترین ضریب نش-ساتکلیف $0/963$ در مرحله صحت‌سنجی دارا می‌باشد.

جدول ۸- نتایج نهایی سه مدل ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین

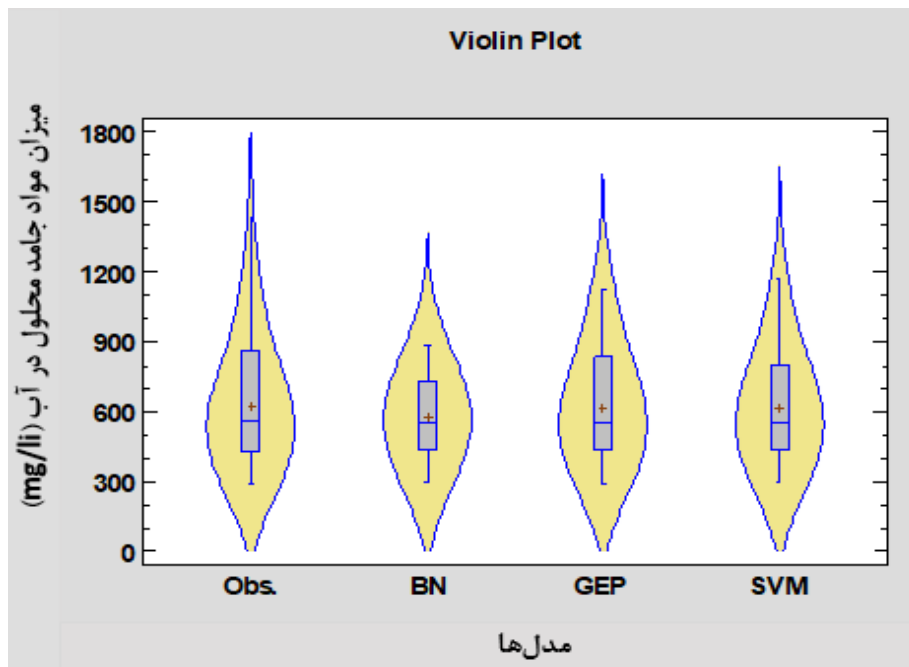
صحت‌سنجی				آموزش				مدل
RMSE	NS	Bias	R^2	RMSE	NS	Bias	R^2	
0/0320	0/963	0/001	0/9815	0/084	0/9505	0/007	0/967	ماشین بردار پشتیبان (هسته پایه شعاعی)
0/0334	0/9598	0/0011	0/9703	0/091	0/933	0/0083	0/948	برنامه‌ریزی بیان ژن
0/092	0/707	0/008	0/8031	0/1229	0/792	0/0151	0/796	شبکه بیزین

در شکل ۶، نتایج بهینه به‌دست آمده از مدل نسبت به زمان نشان داده شده است. با توجه به شکل ۶، مشاهده می‌شود که ماشین بردار پشتیبان در برآورد مقادیر کمینه و بیشینه نسبت به مدل برنامه‌ریزی بیان ژن و شبکه بیزین عملکرد بهتری داشته و توانسته است مقادیر تخمینی را نزدیک به مقادیر مشاهداتی پیش‌بینی نماید. در مقابل، شبکه بیزین در تخمین مقادیر بیشینه عملکرد خوبی نداشته و نتوانسته است دقت مناسبی را ارائه دهد. در تبیین نتایج بالا می‌توان بیان نمود که ماشین بردار پشتیبان برخلاف سایر شبکه‌های عصبی که خطای طبقه‌بندی را کمینه می‌کنند، ریسک عملیاتی را به‌عنوان تابع هدف در نظر گرفته و مقدار بهینه آن را حساب می‌کند. بنابراین، با به‌کارگیری آن با استفاده از توابع پایه شعاعی در پیش‌بینی، باعث می‌شود تا تخمین پارامتر، دقت بالاتر و خطای کمتری نسبت به سایر هسته‌ها داشته باشد. نتایج این تحقیق، با نتایج Dehghani و همکاران (۲۰۱۷) و نتایج Solgi و همکاران (۲۰۱۸) مبنی بر عملکرد بهتر مدل ماشین بردار پشتیبان نسبت به برنامه‌ریزی بیان ژن و شبکه بیزین، هم‌خوانی دارد. در شکل ۷، نمودار ویولن مقادیر بهینه هر سه مدل نسبت به مقادیر مشاهداتی ارائه شده است. همان‌طور که در شکل ۷ مشاهده

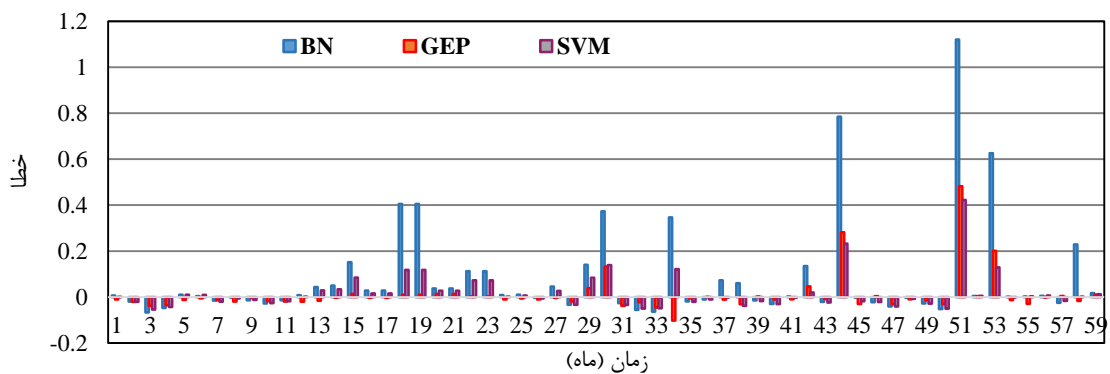
می‌شود، نمودار ویولن مربوط به داده‌های مشاهداتی، مقادیر کمینه و بیشینه به‌ترتیب تقریباً برابر با ۳۰۰ و ۱۴۰۰ دارد که از نظر برآورد نقاط کمینه همه مدل‌ها تقریباً یکسان بوده، اما از نظر نقاط بیشینه مدل ماشین بردار پشتیبان توانسته است مقادیر بیشینه را نزدیک به مقادیر مشاهداتی برآورد کند و بعد از آن مدل برنامه‌ریزی بیان ژن و در آخر شبکه بیزین به مقادیر مشاهداتی تقارن دارند. همچنین، از نمودار استنباط می‌شود که داده‌های مشاهداتی به لحاظ پراکندگی و درصد احتمال با داده‌های محاسباتی از روش‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن، هم‌خوانی بیشتری را نشان می‌دهد که نشان‌دهنده کارایی مناسب مدل‌های یاد شده در برآورد TDS و نیز عملکرد مناسب ماشین بردار پشتیبان می‌باشد. در پایان نیز اختلاف مقادیر مشاهداتی میزان مواد جامد محلول در آب و محاسباتی بهینه مدل‌ها به‌صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به نمونه داده‌های ثبت شده دوره آماری ترسیم شد (شکل ۸). همان‌طور که در این شکل دیده می‌شود، از بین مدل‌های مذکور، مدل ماشین بردار پشتیبان با هسته پایه شعاعی کمترین میزان خطا را داشته است.



شکل ۶- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین در مرحله صحت‌سنجی



شکل ۷- نمودار ویولن هر سه مدل در مرحله صحت‌سنجی



شکل ۸- نمودار خطای بهینه هر سه مدل به صورت درصدی از میانگین مقادیر مشاهداتی

نتیجه‌گیری

در برآورد متغیر مورد نظر مورد استفاده قرار گیرند. همچنین، نتایج حاصل از معیارهای ارزیابی نشان داد که مدل ماشین بردار پشتیبان دقت بالا و خطای کمتری نسبت به مدل‌های برنامه‌ریزی بیان ژن و شبکه بیزین از خود نشان داده است که در تبیین آن می‌توان اظهار نمود که دقت بالای ماشین بردار پشتیبان ناشی از استفاده از اصل استقرای کمینه‌سازی خطای ساختاری است که با به‌کارگیری روش یادگیری با نظارت در توابع پایه شعاعی باعث می‌شود تا تخمین پارامتر از سرعت بالا و خطای کمتری نسبت به سایر مدل‌ها برخوردار شود و این یکی از ویژگی‌های ممتاز توابع پایه شعاعی محسوب می‌شود که با نتایج Asadollahfardi و همکاران (۲۰۱۶) و Haghiahi و همکاران (۲۰۱۸) همخوانی دارد. علاوه بر آن، روند تغییرات میزان مواد جامد محلول در آب نشان می‌دهد که میزان آن با گذشت زمان در حال افزایش بوده که این امر باعث کاهش کیفیت منابع آب سطحی و همچنین، کاهش ظرفیت عبوری جریان رودخانه شده و می‌تواند باعث بروز مشکلاتی در آینده شود. در مجموع، نتایج حاصل از این پژوهش نشان داد که استفاده از روش‌های هوشمند به‌خصوص مدل ماشین بردار پشتیبان می‌تواند در زمینه برآورد میزان مواد جامد محلول در آب موثر باشد.

در این پژوهش، عملکرد مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین جهت برآورد میزان مواد جامد محلول در آب رودخانه با استفاده از داده‌های ماهانه ایستگاه پلدختر واقع در استان لرستان طی سال‌های ۱۳۷۰ تا ۱۳۹۵ مورد بررسی قرار گرفت. مقادیر مواد جامد محلول در آب پیش‌بینی شده در مدل‌های مذکور، بر اساس معیارهای ارزیابی و مواد جامد محلول در آب مشاهداتی مورد مقایسه قرار گرفت. نتایج به‌دست‌آمده از اجرای مدل‌ها با ساختارهای ورودی مختلف نشان داد که در هر سه مدل مورد بررسی، ساختارهای ترکیبی نسبت به سایر ساختارها از عملکرد بهتری برخوردار می‌باشند که نشان‌دهنده انتخاب مناسب پارامترهای ورودی و تاثیرگذار بودن آن‌ها در مدل‌سازی بوده است که این نتایج با نتایج Zeinalzadeh و Zamanzad Ghavidel (۲۰۱۶) مطابقت دارد. از طرفی، افزایش تعداد متغیرهای مؤثر در ورودی موجب گسترش حافظه مدل جهت تخمین مقادیر خروجی می‌باشد که تعداد داده‌ها جهت آموزش شبکه بیشتر شده و شبکه به‌خوبی تعمیم داده شود. قابل ذکر است که در خصوص ترکیب‌هایی با متغیرهای کمتر نیز مدل‌ها جواب‌های خوبی ارائه نمودند که می‌تواند در ایستگاه‌هایی با نواقص آماری

منابع مورد استفاده

1. Aguilera, P.A., A. Fernández, R. Fernández, R. Rumí and A. Salmerón. 2011. Bayesian networks in environmental modelling. *Environmental Modelling and Software*, 26:1376-1388.
2. Ahmadi, F., F. Radmanesh and R. Mirabaisi Najaf Abadi. 2016. Comparing the performance of support vector machines and bayesian networks in predicting daily river flow, case study: Barandoozchay River. *Journal of Water and Soil Conservation*, 22: 171-186 (in Persian).
3. Asadollahfardi, G., A. Meshkat-Dini, Sh. Homayoun Aria and N. Roohani. 2016. Application of artificial neural networks to predict total dissolved solids in the river Zayanderud. *Environmental Engineering Research*, 21: 333-340.
4. Basak, D., S. Pal and D.C. Patranabis. 2007. Support vector regression. *Neural Information Processing, Letters and Reviews*, 11: 203-225.
5. Bozorg-Haddad, O., S. Soleimani and H.A. Loáiciga. 2017. Modeling water-quality parameters using genetic algorithm, least squares support vector regression and genetic programming. *Journal of Environmental Engineering*, 143: 15-32.
6. Cruz-Ramírez, N., H.G. Acosta-Mesa, H. Carrillo-Calvet, L.A. Nava-Fernández and R.E. Barrientos-Martínez. 2007. Diagnosis of breast cancer using bayesian networks. *Computers in Biology and Medicine*, 37: 1553-1564.
7. Dehghani, R., A. Pourhaghi and H. Torabi. 2016. Fluctuations in groundwater level prediction with bayesian network, case study: Noorabad Plain, Lorestan. *Journal of Geographic Space*, 16: 185-200 (in Persian).

8. Dehghani, R., H. Yonesi and H. Torabi Podeh. 2017. Comparing the performance of support vector machines, gene expression programming and bayesian networks in predicting river flow, case study: Kashkan River. *Journal of Water and Soil Conservation*, 24: 161-177 (in Persian).
9. Eskandari, A., R. Noori, H. Meeraji and A. Kiaghaderi. 2012. Developing a proper model for online estimation of the 5-day biochemical oxygen demand based on artificial neural network and support vector machine. *Journal of Environmental Studies*, 38: 71-82 (in Persian).
10. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, 13: 87-129.
11. Ghorbani, M.A., A. Arzani and L. Naghipour. 2016. Comparison of the performance of support vector machine with other intelligent techniques to simulate rainfall-runoff process. *Journal of Watershed Management Research*, 7: 103-92 (in Persian).
12. Ghorbani, M.A. and R. Dehghani. 2016. Application of bayesian neural networks, support vector machines and gene expression programming analysis of rainfall-runoff monthly, case study: Kakareza River. *Irrigation Sciences and Engineering*, 39: 125-138 (in Persian).
13. Ghorbani, M.A. and R. Dehghani. 2017. Comparison of bayesian neural network, artificial neural network gene expression programming in river water quality, case study: Belkhviachay River. *Journal of Watershed Management Research*, 8: 13-24 (in Persian).
14. Haghiabi, A.H., A.H. Nasrolahi and A. Parsaie. 2018. Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53: 3-13.
15. Hamel, L.H. 2011. Knowledge discovery with support vector machines. John Wiley and Sons, 256 pages.
16. Hosseini, P., A.R. Ildoromi and A. R. Hosseini. 2013. Water quality assessment of Karun River using NSFQI index in Zargan to Kut Amir (over 5 years) . *Journal of Environmental Sciences and Technology*, 11: 1-11 (in Persian).
17. Kanda, E.K., E.C. Kipkorir and J.R. Kosgei. 2016. Dissolved oxygen modelling using artificial neural network: a case of River Nzoia, Lake Victoria basin, Kenya. *Journal of Water Security*, 2: 1-8.
18. Khaled, B., A. Abdellah, D. Noureddine, H. Salim and A. Sabeha. 2018. Modelling of biochemical oxygen demand from limited water quality variable by ANFIS using two partition methods. *Water Quality Research Journal*, 53: 24-40.
19. Kuikka, S. and O. Varis. 1997. Uncertainties of climatic change impacts in Finnish Watersheds: a bayesian network analysis of expert knowledge. *Boreal Environment Research*, 2: 109-128.
20. Lerner, U., R. Parr, D. Koller and G. Biswas. 2000. Bayesian fault detection and diagnosis in dynamic systems. *The Association for the Advancement of Artificial Intelligence (AAAI)/ Innovative Applications of Artificial Intelligence (AAAI/IAAI)*, 531-537.
21. Liu, G.Q. 2011. Comparison of regression and ARIMA models with neural network models to forecast the daily stream flow of White Clay Creek. PhD Thesis, University of Delaware, 545 pages.
22. Montaseri, M. and S. Zamanzad Ghavidel. 2016. Comparing the performance of artificial intelligence models in estimating water quality parameters in periods of low and high water flow. *Journal of Water and Soil*, 30: 1733 -1747 (in Persian).
23. Muangthong, S. 2015. Assessment of surface quality using multivariate statistical techniques, a case of the Nampong River basin, Thailand. *The Journal of Industrial Technology*, 11: 25-37.
24. Rezaei, E., A. Khashei-Siuki and A. Shahidi. 2014. Design of Groundwater level monitoring network, using the model of Least Squares Support Vector Machine (LS-SVM). *Iranian Journal of Soil and Water Research*, 45: 389-396 (in Persian).
25. Sadeghi Hesar, A., H. Tabatabaee and M. Jalali. 2012. Monthly rainfall forecasting using bayesian belief networks. *International Research Journal of Applied and Basic Sciences*, 3: 2226-223130.
26. Sattari, M.T., A.R. Joudi and A. Kusiak. 2016. Estimation of water quality parameters with data-driven model. *Journal-American Water Works Association*, 108: 232-239.
27. Singh, K.P., A. Basant, A. Malik and G. Jain. 2009. Artificial neural network modeling of the river water quality, a case study. *Journal of Ecological Modeling*, 220: 888-895.
28. Solgi, A., H. Zarei, M. Shani Darabi and S. Alidadi Deh Kohneh. 2018. Application of gene expression programming and support vector regression models to modeling and prediction monthly precipitation. *Journal of Applied Researches in Geographical Sciences*, 18: 91-103 (in Persian).
29. Xu, L., J. Wang, J. Guan and F. Huang. 2007. A support vector machine model for mapping of lake water quality from remote-sensed images. *International Journal of Intelligent Computing in Medical Sciences and Image Processing*, 1: 57-66.
30. Xu, H., H. Zheng, X. Chen, Y. Ren and Z. Ouyang. 2016. Relationships between river water quality and landscape factors in Haihe River Basin, China: implications for environmental management. *Chinese Geographical Science*, 26: 197-207.

31. Zamanzad Ghavidel, S. and K. Zeinalzadeh. 2016. Estimation of rivers dissolved solids TDS by soft computing, case study: upstream of Boukan Dam. Journal of Water and Soil, 29: 1262-1277 (in Persian).

Assessment of intelligent models in estimating the Total Dissolved Solids in Kashkan River, Lorestan Province

Elham Rezaei¹, Babak Shahinejad^{*2} and Hojatollah Younesi³

¹ MSc Student, Faculty of Agriculture, Lorestan University, Khorramabad, Iran ^{2 and 3} Assistant Professor, Faculty of Agriculture, Lorestan University, Khorramabad, Iran

Received: 29 July 2018

Accepted: 18 September 2018

Abstract

One of the important issues concerning the rivers quality is prediction of Total Dissolved Solids (TDS) in water. In this study, the performance of Support Vector Machine (SVM) intelligent models with different kernel functions, Gene Expression Programming (GEP) and Bayesian Networks (BN) were investigated in estimating the TDS in Kashkan River water. For this purpose, the quality data of Poldokhtar Station located in Lorestan Province were utilized to predict the TDS in water during the statistical period of 1991-2016 which included the hydrogen carbonate, chloride, sulfate, magnesium, calcium, sodium, electrical conductivity, flow rate and pH. To validate the models the criteria such as the coefficient of determination (R^2), Nash-Sutcliffe coefficient, Root Mean Square Error (RMSE), and bias coefficient were applied. The obtained results showed that in all three models, the combinational structures have acceptable accuracy. Also, according to the assessment criteria, it was found that the SVM with Radial Basis Function (RBF) kernel has the highest accuracy of 0.982, minimum root square error of 0.032 *mg/lit*, minimum bias of 0.001 and highest Nash-Sutcliffe coefficient value of 0.963 with respect to other models.

Keywords: Bayesian Network, Different kernel functions, Forecast, Gene Expression Programming, Support Vector Machine

* Corresponding author: shahinejad.b@lu.ac.ir