

Comparison of Wavelet Neural Network Models, Support Vector Machine, and Gene Expression Programming in Estimating the amount of Oxygen Dissolved in Rivers

B. Shahinejad^{1*} and R. Dehaghani²

Abstract

Dissolved oxygen in rivers is one of the most effective parameters in determining water quality and its control is one of the most important factors in development of water resources in each region. For this reason, we investigated the performance of wavelet neural network models, support vector machines, and gene expression programming for estimating the dissolved oxygen in Cumberland River in Tennessee. The indicators of the Cumberland River, including DO, flow rate and temperature during 10-year period (2006-2016), were simulated in monthly time scale. Also for evaluation and performance of the models, the correlation coefficient, root mean square error and mean absolute magnitude error were applied. The results showed that integrated input structures into models in all three models offer better performance than other structures. Also, the results of the evaluation criteria showed that the wavelet neural network model showed the highest correlation coefficient (0.960), lowest root mean square error (0.668), and the lowest mean error magnitude (519.0). In general, the results showed that due to the high ability of the wavelet neural network and the elimination of time series noise in the estimation of river water quality parameters, this model could be a suitable and rapid solution for water resource quality management.

Keywords: Dissolved Oxygen, Gene Expression Programming, Wavelet Neural Network, Water Quality, Support Vector Machine.

Received: February 22, 2018

Accepted: April 20, 2018

مقایسه مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در تخمین میزان اکسیژن محلول در آب رودخانه‌ها

بابک شاهی‌نژاد^{۱*} و رضا دهقانی^۲

چکیده

اکسیژن محلول در آب رودخانه‌ها از مؤثرترین پارامترها در تعیین کیفیت آب و کنترل آن از مهم‌ترین عوامل توسعه منابع آب هر منطقه است. به همین دلیل در این پژوهش عملکرد مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن به منظور تخمین اکسیژن محلول در آب رودخانه کامبرلند واقع در ایالت تنسی مورد بررسی قرار گرفت. برای این منظور جهت شبیه‌سازی از شاخص‌های موجود رودخانه کامبرلند شامل DO، دبی جریان و دما در طی یک دوره ۱۰ ساله (۲۰۰۶-۲۰۱۶) در مقیاس زمانی ماهانه استفاده شد. همچنین، جهت ارزیابی و عملکرد مدل‌ها معیارهای ضریب همبستگی، ریشه میانگین مربعات خطا و میانگین قدر مطلق خطا بکار گرفته شد. نتایج نشان داد ساختارهای ترکیبی ورودی به مدل‌ها، در هر سه مدل عملکرد بهتری نسبت به سایر ساختارها ارائه می‌دهند. علاوه بر آن، نتایج حاصل از معیارهای ارزیابی نشان داد از بین مدل‌های بکار رفته، مدل شبکه عصبی موجک بیشترین ضریب همبستگی (۰/۹۶۰)، کمترین جذر میانگین مربعات خطا (۰/۶۶۸) و نیز کمترین میانگین قدر مطلق خطا (۵۱۹/۰) را از خود نشان داده است. در مجموع نتایج نشان داد به لحاظ توانایی بالای شبکه عصبی موجک و حذف نویزهای سری‌های زمانی در تخمین پارامترهای کیفی آب رودخانه، این مدل می‌تواند، راهکاری مناسب و سریع در مدیریت کیفیت منابع آب مطرح شود.

کلمات کلیدی: اکسیژن محلول در آب، برنامه‌ریزی بیان ژن، شبکه عصبی موجک، کیفیت آب، ماشین بردار پشتیبان.

تاریخ دریافت مقاله: ۹۶/۱۲/۳

تاریخ پذیرش مقاله: ۹۷/۱/۳۱

1- Assistant Professor of Water Engineering University of Lorestan, Khorramabad, Iran. Email: shahinejad.b@lu.ac.ir

2 Ph.D. Student of Water Structure, Faculty of Agriculture, University of Lorestan, Khorramabad, Iran.

*- Corresponding Author

۱- استادیار گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان.

۲- دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی، دانشگاه لرستان.

*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان زمستان ۱۳۹۷ امکانپذیر است.

توسعه روزافزون فعالیت‌های کشاورزی و صنعتی و افزایش حجم فاضلاب‌های شهری موجب آلودگی رودخانه‌ها شده است، به نحوی که کیفیت این منابع حیاتی را مورد مخاطره جدی قرار داده است. همچنین مصرف بی‌رویه مواد شیمیایی در کشاورزی، باعث افزایش شدت آلودگی منابع آب شهرها و روستاها که به اندازه کافی آلوده شده‌اند، می‌شود (Ghorbani and Dehghani, 2017). DO، از مهم‌ترین شاخص‌های کیفی برای بررسی سلامت رودخانه‌ها می‌باشد (Dogan et al., 2009). DO، مقدار اکسیژن محلول در آب به عنوان یک فاکتور مهم و موثر در زندگی جانداران آبی و آلودگی آب می‌باشد (Chapman, 1992). همچنین میزان زیاد DO سبب ایجاد شرایط نامناسب زندگی برای گیاهان و جانوران رودخانه می‌شود (Radwan et al., 2003). امروزه برای اندازه‌گیری اکسیژن محلول در محل نمونه‌برداری از دستگاه‌های کوچک سیار که مجهز به الکتروود غشایی است، استفاده می‌شود. الکتروود غشایی بر اساس سرعت نفوذ مولکول‌های اکسیژن از یک غشا ساخته شده است. این روش فیزیکی بطور ساده و سریع انجام می‌گیرد. از طرفی دیگر دقیق‌ترین روش اندازه‌گیری اکسیژن محلول روش یدومتری است. این روش یک روش تیتراسیون است که بر اساس خواص اکسیدکنندگی اکسیژن محلول انجام می‌گیرد.

از آنجایی که کیفیت آب رودخانه‌ها از مشخصه‌های مختلفی که رفتاری پیچیده و غیرخطی دارند تأثیر می‌پذیرند مدل‌های ریاضی ممکن است عملکرد خوبی نداشته باشند (Dogan et al., 2009). امروزه سیستم‌های هوشمند به‌طور گسترده برای پیش‌بینی پدیده‌های غیرخطی مورد استفاده قرار می‌گیرد، که روش شبکه عصبی موجک (WNN^1) ، برنامه‌ریزی بیان ژن (GEP^2) و ماشین بردار پشتیبان (SVM^3) از جمله‌ی این روش‌ها به‌شمار می‌روند. برنامه‌ریزی بیان ژن یک تکنیک برنامه‌ریزی خودکار است که راه‌حل مسأله را با استفاده از برنامه‌ریزی کامپیوتر ارائه کرده و عضوی از خانواده الگوریتم تکاملی می‌باشد (Ghorbani and Salehi, 2011). مدل ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است (Dehghani et al., 2015). در سال‌های اخیر استفاده از روش‌های هوشمند در مطالعات کیفی رودخانه‌ها مورد توجه محققین قرار گرفته است، که از جمله می‌توان به موارد ذیل اشاره نمود:

Eskandari et al. (2012) عملکرد دو روش ماشین بردار پشتیبان و شبکه عصبی مصنوعی را در پیش‌بینی اکسیژن خواهی بیوشیمیایی ۵ روزه رودخانه سفیدرود مورد ارزیابی قرار دادند نتایج مطالعه آن‌ها

نشان داد از بین دو روش مذکور، مدل ماشین بردار پشتیبان با دقت بیشتری شاخص‌های کیفی رودخانه را شبیه‌سازی می‌کند. Banejad et al. (2013) جهت تخمین برخی پارامترهای کیفی رودخانه‌های جاجرود و قره‌سو طی بازه زمانی ۱۳۸۳-۱۳۶۰ از مدل شبکه عصبی موجک استفاده کردند و به این نتیجه رسیدند که مدل شبکه عصبی موجک توانایی قابل‌قبولی نسبت به شبکه عصبی مصنوعی در تخمین پارامترهای کیفی رودخانه دارد. Radmanesh et al. (2016) به مقایسه عملکرد مدل‌های شبکه عصبی مصنوعی و شبکه عصبی مصنوعی موجکی جهت پیش‌بینی اکسیژن مورد نیاز بیولوژیکی در ایستگاه ملاثانی پرداختند. نتایج آن‌ها نشان داد اعمال تبدیل موجک بر روی ورودی داده‌ها که منجر به شبکه عصبی موجکی می‌شود، از میزان خطای کمتری برخوردار است. Ghorbani et al. (2015) از مدل‌های هوشمند جهت پیش‌بینی فرآیند بارش رواناب حوضه بالیخوچای استفاده کردند نتایج نشان داد مدل هیبرید موجک- شبکه عصبی برای پیش‌بینی رواناب در مقیاس زمانی روزانه از دقت بالایی برخوردار است. Nozari and Tavakoli (2017) از مدل‌های هوشمند سری زمانی و داده‌های حوضه دریاچه ارومیه در طی یک دوره ۱۵ ساله جهت تولید و بازسازی جریان استفاده کردند نتایج نشان داد مدل شبکه عصبی مصنوعی در حفظ توأم خواص ماهانه جریان عملکرد قابل‌قبولی دارد. Solgi et al. (2017) برای پیش‌بینی اکسیژن مورد نیاز بیولوژیکی از ترکیب مدل ماشین بردار پشتیبان با تبدیل موجک استفاده کردند نتایج نشان داد مدل ترکیبی از عملکرد بهتری نسبت به مدل ماشین بردار پشتیبان برخوردار است. Chu et al. (2013) برای پیش‌بینی اکسیژن بیولوژیکی محلول در آب رودخانه لیاو شرقی چین از شبکه عصبی مصنوعی استفاده کردند نتایج آن‌ها نشان داد شبکه عصبی مصنوعی دقت قابل‌قبولی در تخمین شاخص‌های کیفی رودخانه دارد. Alizadeh et al. (2015) استفاده از مدل ترکیبی شبکه عصبی مصنوعی موجک را برای پیش‌بینی پارامترهای کیفی در خلیج هیلو^۴، اقیانوس آرام انجام دادند و به این نتیجه دست یافتند که مدل ترکیبی شبکه عصبی مصنوعی موجک عملکرد بهتری نسبت به شبکه عصبی مصنوعی دارد. نتایج حاصل از پژوهش Ghorbani et al. (2016) که جهت پیش‌بینی کیفیت آب رودخانه زرینه‌رود از مدل‌های شبکه عصبی مصنوعی، توابع پایه شعاعی و ماشین بردار پشتیبان استفاده کردند نشان داد مدل ماشین بردار پشتیبان از توانایی بالایی در پیش‌بینی کیفیت آب رودخانه برخوردار است. نتایج حاصل از پژوهش Kishorekumar and Velrajan. (2017) جهت پیش‌بینی کیفیت آب دریاچه کودای^۵ واقع در چین از مدل برنامه‌ریزی بیان ژن استفاده نمودند نشان داد مدل برنامه‌ریزی بیان ژن با ساختاری متشکل از چهار عملگر اصلی ریاضی در پیش‌بینی کیفیت آب از دقت بالایی

همچنین رودخانه کامبرلند از شهرستان لچر در کنتاکی شرقی در دشت کامبرلند آغاز شده و به رودخانه اوهایو در اسمیثلند می‌ریزد. این رودخانه در بین طول‌های جغرافیایی " ۵۷' ۵۱° ۸۲ درجه تا " ۲۲' ۳۶° ۸۴ شرقی و عرض‌های جغرافیایی " ۵۰' ۷' ۳۷ تا " ۲۸' ۵۲° ۳۸ درجه شمالی در ایالت تنسی واقع شده است. در شکل ۱ موقعیت ایستگاه مورد مطالعه در طول جغرافیایی " ۵۶' ۴۹° ۸۶ درجه و عرض جغرافیایی " ۵۹' ۱۰° ۳۶ نشان داده شده است. در این پژوهش جهت تخمین میزان اکسیژن محلول در آب از داده‌های ماهانه اکسیژن محلول در آب (میلی گرم در لیتر)، بارش (میلی متر)، دبی جریان (مترمکعب بر ثانیه) و دما (درجه سانتی‌گراد) مربوط به ایستگاه کامبرلند در طی سال آماری ۲۰۰۶-۲۰۱۶ که از سازمان زمین‌شناسی ایالات متحده اخذ شد، استفاده گردید.

۲-۲- برنامه‌ریزی بیان ژن

برنامه‌ریزی بیان ژن تخمین‌یافته الگوریتم ژنتیک می‌باشد که برای اولین بار بر اساس تئوری داروین ارائه شد. به این ترتیب که جمعیتی در جهت تکامل به صورت انتخابی، جمعیت نامناسب را رها کرده و فرزندان اصلاح‌شده ایجاد می‌کنند. برنامه‌ریزی بیان ژن یک تکنیک برنامه‌ریزی خودکار می‌باشد که راه‌حل مسأله را با استفاده از برنامه کامپیوتری ارائه می‌کند. در این روش در ابتدای فرآیند هیچ‌گونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مؤلفه‌های آن می‌باشد. برنامه‌ریزی بیان ژن برخلاف الگوریتم ژنتیک روی ساختار درختی فرمول‌ها به جای سلسله ارقام دودویی عمل می‌کند. ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسأله و اعداد ثابت) ایجاد می‌شوند (Ferreira, 2001).

برخوردار است. (Mehdipour et al. (2017 در پژوهشی کاربرد مدل برنامه‌ریزی بیان ژن را جهت تخمین میزان اکسیژن محلول در آب رودخانه‌ها با استفاده از آمار و اطلاعات ۴ ایستگاه هیدرومتری دریاچه ایمیر^۱ واقع در ترکیه بررسی کردند. نتایج پژوهش آن‌ها بیانگر عملکرد قابل‌قبول مدل برنامه‌ریزی بیان ژن در تخمین این شاخص کیفی می‌باشد. (Raheli et al. (2017 برای پیش‌بینی اکسیژن بیولوژیکی محلول در آب رودخانه لانگات^۲ واقع در مالزی از مدل ترکیبی شبکه عصبی با الگوریتم کرم شب‌تاب استفاده کردند. نتایج آن‌ها حاکی از دقت و توانایی بالای مدل پیشنهادی برای پیش‌بینی و شبیه‌سازی پارامتر اکسیژن محلول در آب بود. از سوی دیگر کاهش کیفیت آب رودخانه کامبرلند که به‌عنوان مهم‌ترین منبع تأمین‌کننده آب بخش‌های مختلف نواحی مجاور خود از جمله ایالت تنسی، به دلیل وجود مواد شیمیایی احیاکننده و به‌ویژه مواد آلی و نیز تخلیه فاضلاب‌ها و پساب‌های صنعتی و خانگی سبب بروز مشکلات زیادی شده است. بنابراین با توجه به پژوهش‌های انجام‌شده و ضرورت مدل‌سازی کیفی آب این رودخانه جهت بهبود کیفیت در این تحقیق سعی گردید جهت تخمین اکسیژن محلول در آب ایستگاه کامبرلند واقع در ایالت تنسی از مدل شبکه‌های عصبی موجک، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان بر اساس متغیرهایی اندازه‌گیری شده موجود در ایستگاه کامبرلند همچون اکسیژن محلول در آب، دبی جریان و دما در مقیاس زمانی ماهانه استفاده گردد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

رودخانه کامبرلند از رودخانه‌های مهم جنوب ایالات متحده بوده که طول آن ۱۱۰۶ کیلومتر است. این رودخانه در حوزه آبریز جنوب ایالت کنتاکی^۳ در مساحتی بالغ بر ۴۷۰۰ کیلومترمربع واقع شده است.

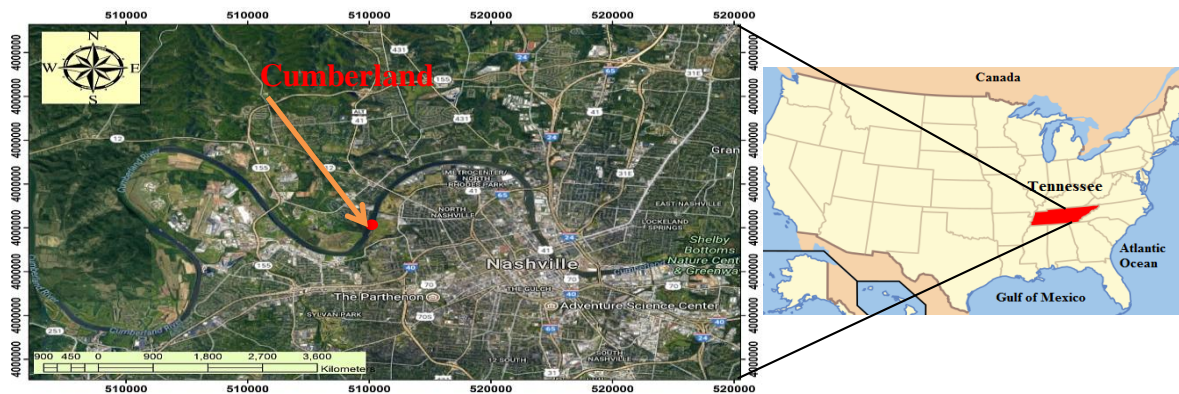


Fig. 1- The study area
شکل ۱- منطقه مورد مطالعه

بردارای متعامد با v_m در رزولوشن m فرض کنیم آنگاه می توان دسته های دیگر از پایه های متعامد فضای wm که ویولت $\Psi(x)$ نامیده می شوند را نیز به صورت زیر بیان نمود (Wang et al., 2000):

$$\Psi_{m,k}(x) = 2^{-m/2} \Psi(2^{-m}x - k) \quad m, k \in \mathbb{Z} \quad (2)$$

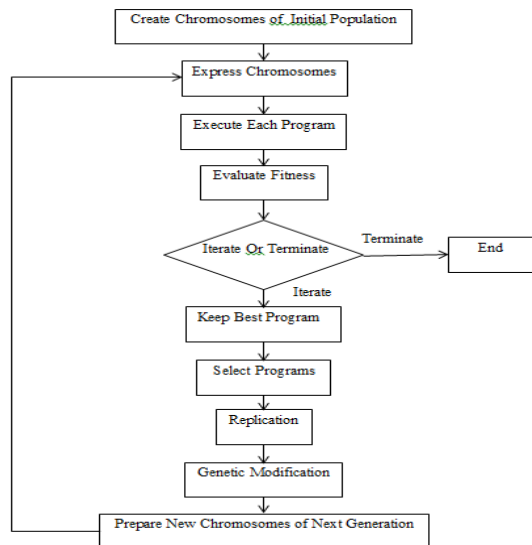


Fig. 2- The process of implementing gene expression programming

شکل ۲- فرآیند اجرای برنامه ریزی ژنتیک

در حالت کلی همه توابع فیزیکی را می توان به کمک ویولت ها و توابع مقیاس به صورت زیر بیان نمود:

$$f(x) = f_0(x) + \sum_{m=-\infty}^0 \sum_{k=-\infty}^{\infty} d_{m,k} \Psi_{m,k} \quad (3)$$

$$f_0(x) = \sum_k a_{0,k} \Phi_{0,k} \quad (4)$$

روابط فوق بیان می دارند که هر تابع فیزیکی را می توان ابتدا به کمک تعدادی توابع مقیاس در رزولوشن صفر تقریب زد و سپس به کمک توابع ویولت در رزولوشن های مختلف آن تقریب را تا میزان دقت مورد نظر ادامه داد.

شبکه عصبی ویولت بر مبنای روابط (۳) و (۴) شکل می گیرد و دارای الگوریتم های آموزش مخصوص خود می باشد که در پژوهش Shin et al. (2005) به صورت کامل مورد بررسی قرار گرفته است. لازم به ذکر است در حالت کلی خانواده ویولت های پیوسته نیز به صورت زیر بیان می شود:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R} \quad (5)$$

و تبدیل ویولت برای توابع پیوسته به وسیله رابطه زیر محاسبه می شود:

فرآیند اجرایی گام به گام برنامه ریزی بیان ژن به صورت مراحل زیر است: ۱. تولید یک جمعیت اولیه از فرمول ها که این فرمول ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول ها) و ترمینال ها (متغیرهای مسأله و اعداد ثابت) ایجاد می شوند، ۲. هر یک از افراد جمعیت مذکور با استفاده از توابع برازش مورد ارزیابی قرار می گیرند، ۳. تولید یک جمعیت جدید از فرمول ها، که مراحل زیر برای تولید یک جمعیت جدید دنبال می شود:

الف. یکی از عمل های ژنتیکی تلاقی^۱، جهش^۲ و تولید مثل^{۱۱} انتخاب می شود (این سه عمل ژنتیکی، مهم ترین عمل های ژنتیکی مورد استفاده در برنامه ریزی ژنتیک می باشند. عمل های دیگری مثل اصلاح ساختار و غیره نیز با احتمال کمتر مورد استفاده قرار می گیرند). ب. تعداد مناسبی از افراد جمعیت حاضر انتخاب می شوند (انتخاب فرد یا افرادی از جمعیت مذکور به صورت احتمالاتی می باشد که در این انتخاب احتمالاتی منفردهای با برازش بهتر به منفردهای نامرغوب ترجیح داده می شوند و این بدان معنی نیست که حتماً منفردهای نامرغوب حذف می شوند). ج. از عمل ژنتیکی انتخاب شده برای تولید فرزند (فرمول جدید) استفاده می شود. د. فرزند (فرمول جدید) تولید شده در یک جمعیت جدید وارد می شود و ه. مدل مورد نظر با استفاده از تابع برازش مورد ارزیابی واقع می شود. ۴. گام سوم تا نیل به حداکثر تعداد تولید، تکرار خواهد شد.

۲-۳- شبکه عصبی موجک

شبکه های عصبی مبتنی بر ویولت^{۱۲} که ویولت^{۱۳} نامیده می شوند از ترکیب دو تئوری ویولت ها و شبکه های عصبی به وجود آمده اند (Vapnik, 1998). این شبکه ها هم از مزایا و ویژگی های شبکه های عصبی برخوردارند و هم جذابیت ها و انعطاف پذیری ها و مبانی ریاضی مستحکم ویولت ها و آنالیز چند مقیاسی را دارا می باشند. در ویولت ها از دو دسته توابع ویولت و توابع مقیاس (scaling functions) استفاده می شود. یک خانواده از توابع مقیاس $\Phi(x)$ به صورت زیر معرفی می شوند:

$$\Phi_{m,k}(x) = 2^{-m/2} \Phi(2^{-m}x - k) \quad m, k \in \mathbb{Z} \quad (1)$$

اگر $\Phi(x)$ را یک تابع مقیاس مادر در نظر بگیریم، مقادیر 2^{-m} و k به ترتیب مربوط به گسترش و انتقال تابع مقیاس مادر هستند. توابع مقیاس رزولوشن m ، $\Phi_{m,k}(x)$ ، در اصل پایه های متعامد فضای برداری v_m در رزولوشن m می باشند. به عبارتی فضای برداری v_m حاوی کلیه تقریب های تابع $f(x)$ با رزولوشن m و توسط توابع $\Phi_{m,k}(x)$ می باشد. بنابراین فضاهای برداری $\{v_m\}$ حاوی تقریب های مختلف تابع $f(x)$ در رزولوشن های مختلف می باشند. حال اگر wm را فضای

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \phi(x_i)^T \cdot \phi(x) + b \quad (12)$$

در معادله ۱۲، $\bar{\alpha}_i$ میانگین ضرایب لاگرانژ می‌باشد. محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد (Yoon et al., 2011). برای حل این مشکل روند معمول در مدل SVM رگرسیون انتخاب یک تابع کرنل به صورت رابطه زیر است.

$$K(X_j, X) = \phi(X_j)^T \sqrt{b^2 - 4ac} \quad (13)$$

می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف SVM- ϵ استفاده کرد. انواع توابع کرنل قابل استفاده در مدل SVM رگرسیونی شامل کرنل چندجمله‌ای^{۱۴} و کرنل توابع پایه شعاعی^{۱۵} (RBF) و کرنل خطی، به ترتیب طبق روابط ۱۴، ۱۵ و ۱۶ محاسبه می‌گردند. قابل ذکر است فرآیند محاسبات ماشین بردار پشتیبان بر اساس کد نویسی در محیط MATLAB انجام، و پارامترهای توابع کرنل از طریق سعی و خطا بهینه گردیدند.

$$k(x, x_j) = (t + x_i \cdot x_j)^d \quad (14)$$

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (15)$$

$$k(x, x_j) = x_i \cdot x_j \quad (16)$$

۲-۵- معیارهای ارزیابی

در این تحقیق به منظور ارزیابی دقت و کارایی مدل‌ها، از نمایه‌های ضریب همبستگی (R)، جذر میانگین مربعات خطا (RMSE) و میانگین قدر مطلق خطا (MAE) طبق روابط زیر استفاده گردید. بهترین مقدار برای این سه معیار به ترتیب یک، صفر و صفر می‌باشد.

$$R^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad -1 \leq R \leq 1 \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (19)$$

در روابط بالا، x_i و y_i به ترتیب مقادیر مشاهداتی و محاسباتی در گام زمانی i ام، N تعداد گام‌های زمانی، \bar{x} و \bar{y} نیز به ترتیب میانگین مقادیر مشاهداتی و محاسباتی می‌باشد. علاوه بر معیارهای فوق از نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی-محاسباتی نسبت به زمان نیز جهت تحلیل بیشتر نتایج استفاده شده است.

۳- نتایج و بحث

یکی از مهم‌ترین مراحل در مدل‌سازی، انتخاب ترکیب مناسبی از متغیرهای ورودی است. در مدل‌های هوشمند انتخاب ورودی‌های اولیه

$$W_{a,b}(f) = \bar{f}_{(a,b)} = \langle \Psi_{a,b}(t), f(t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \Psi^*\left(\frac{t-b}{a}\right) dt \quad (6)$$

در اینجا پارامتر a به عنوان پارامتر گسترش و متناسب با فرکانس می‌باشد (به عبارت دیگر پارامتر تأخیر کوچک متناظر با فرکانس بالا و پارامتر تأخیر بزرگ متناظر با فرکانس پایین می‌باشد) و پارامتر b به عنوان پارامتر انتقال و متناسب با زمان می‌باشد.

۲-۴- ماشین بردار پشتیبان

ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد (Vapnik, 1995). در مدل رگرسیون SVM تابعی مرتبط با متغیر وابسته Y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ϵ) مشخص شود (Vapnik, 1995):

$$f(x) = W^T \cdot \phi(x) + b \quad (7)$$

$$y = f(x) + \text{noise} \quad (8)$$

چنانچه W^T ترانهاده بردار ضرایب و b ثابت مشخصه‌های تابع رگرسیونی و ϕ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با آموزش مدل SVM توسط مجموعه‌ای از داده‌ها (مجموعه آموزش) محقق می‌شود (Misra et al., 2009). برای محاسبه W و b لازم است تابع خطا (معادله ۹) در مدل SVM- ϵ با در نظر گرفتن شرایط مندرج (قیود) در معادله (۱۰) و (۱۱) کمینه شود (Hamel, 2009):

$$\frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \epsilon_i + C \sum_{i=1}^N \epsilon_i^* \quad (9)$$

$$W^T \cdot \phi(X_i) + b - y_i \leq \epsilon + \epsilon_i^* \quad (10)$$

$$y_i - W^T \cdot \phi(X_i) - b \leq \epsilon + \epsilon_i, \quad \epsilon_i, \epsilon_i^* \geq 0, \quad i=1, 2, \dots, N \quad (11)$$

در معادلات بالا C عددی صحیح و مثبت است، که عامل تعیین جریمه در هنگام رخ دادن خطای آموزش مدل است. ϕ تابع کرنل، N تعداد نمونه‌ها و دو مشخصه ϵ_i و ϵ_i^* متغیرهای کمبود هستند که حد بالا و پایین خطای آموزش مرتبط با مقدار خطای مجاز ϵ را مشخص می‌کند. در مسائل پیش‌بینی می‌شود که داده‌ها، درون بازه مرزی ϵ قرار گیرند حال اگر داده‌ای خارج از بازه ϵ قرار گرفت آنگاه یک خطای معادل ϵ_i و ϵ_i^* وجود خواهد داشت. در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد:

$$X_n = 0.1 + 0.8 \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (20)$$

۳-۱- نتایج شبکه عصبی موجک

به منظور مدل سازی اکسیژن محلول در آب از مدل شبکه عصبی موجکی با لایه های پنهان و با تعداد نرون های متفاوت استفاده شده است. در مدل عصبی موجکی، ابتدا موجک مناسب (سیملت) انتخاب گردید. سپس با اعمال تبدیل روی داده ها، ضرایب تقریب و جزئیات آن ها استخراج گردید و داده ها به وسیله تابع موجک کلاسه مکزیکی به عنوان توابع فعال سازی که مشتق دوم تابع گوسی بوده، تبدیل می یابند. جهت آموزش شبکه، الگوریتم گرادینان نزولی که در یادگیری شبکه های عصبی و حداقل سازی مقدار خطای و تنظیم پارامترهای شبکه استفاده می شود، بکار گرفته شد. همان طور که در جدول ۳ مشاهده می شود ساختار شماره ۴ با ضریب همبستگی ۰/۹۶۰ و کمترین ریشه میانگین مربعات خطا ۰/۶۶۸ میلی گرم بر لیتر و کمترین میانگین قدر مطلق خطا ۰/۵۱۹ میلی گرم بر لیتر در مرحله صحت سنجی نسبت به سایر ساختارها جهت مدل سازی مناسب تر خواهد بود. در شکل ۳ نمودار بهترین مدل حاصل شده برای داده های بخش آموزش و صحت سنجی نشان داده شد. همان طور که از شکل (۳-۵) مشخص است بیشتر مقادیر تخمینی و مشاهداتی به غیر از چند نقطه روی خط نیمساز ($y=x$) قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی است. در شکل (۳-۵) نیز تغییرات مقادیر محاسباتی و مشاهداتی نسبت به زمان نشان داده شد، به طوری که این مدل در تخمین اکثر مقادیر دقت قابل قبولی داشته به گونه ای که این مقادیر را نزدیک به مقدار واقعی شان تخمین زده است.

مناسب و تأثیرگذار در پدیده به منظور آموزش ماهیت سازوکار حاکم بر پدیده باعث بهبود عملکرد خواهد شد. بنابراین در الگوسازی میزان اکسیژن محلول در رودخانه کامبرلند نیز سعی گردید تا مؤثرترین پارامترها جهت آموزش مدل ها انتخاب شود که در جدول ۲ نشان داده شده است (Nourani et al., 2016; Satari et al., 2016). در این پژوهش به منظور تخمین میزان اکسیژن محلول در آب رودخانه از داده های ماهانه ایستگاه کامبرلند در طول دوره آماری ۲۰۰۶ تا ۲۰۱۶ استفاده شد که در جدول ۱ مقادیر داده های مورداستفاده نشان داده شده است. جهت مدل سازی، پارامترهای دبی جریان (Q) و دما (T) در گام های زمانی t و t-1 به عنوان ورودی و میزان اکسیژن محلول در آب (DO) به عنوان پارامتر خروجی مدل ها بکار برده شد. همچنین در جدول ۲ ترکیب های مختلف ورودی نشان داده شد. لازم به ذکر است جهت مدل سازی ۸۰ درصد داده ها برای آموزش و ۲۰ درصد باقی مانده جهت تست، به صورت تصادفی، که گستره وسیعی از انواع داده ها را پوشش دهد، انتخاب شد (Nagy et al., 2002; Kisi and Karahan, 2006). در ادامه نتایج مدل های مورداستفاده ارائه می گردد:

در این میان یک نکته مهم در آموزش شبکه های عصبی نرمال سازی داده ها قبل از استفاده در مدل می باشد این عمل خصوصاً وقتی دامنه تغییرات ورودی ها زیاد باشد کمک شایانی به آموزش بهتر و سریع تر مدل می کند. اصولاً وارد کردن داده ها به صورت خام باعث کاهش سرعت و دقت شبکه می شود (Zhu et al., 2007). برای نرمال سازی داده ها از رابطه ی ۲۰ استفاده شد.

Table 1- The statistical properties of the parameters for the statistical period (2006-2016)
جدول ۱- خصوصیات آماری پارامتر مورد استفاده در طی دوره آماری (۲۰۰۶-۲۰۱۶)

Parameter	Training			Testing		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Q(m ³ /s)	1.203	25.692	156.139	4.295	32.347	70.593
T(°C)	5.33	17.204	28.41	5.62	16.272	26.18
DO(mg/l)	5.53	9.588	14.96	6.03	9.424	13.02

Table 2- Selected compositions of input parameters for the wavelet neural network, support vector machine and gene expression programming models

جدول ۲- ترکیب های منتخب پارامترهای ورودی مدل های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه ریزی بیان ژن

Number	Input	Output
1	Q(t)	DO(t)
2	Q(t), T(t)	DO(t)
3	Q(t), T(t), Q(t-1)	DO(t)
4	Q(t), T(t), Q(t-1), T(t-1)	DO(t)

Table 3- The analysis of wavelet neural network model results for different input structures

جدول ۳- تحلیل نتایج مدل شبکه عصبی موجک برای ساختارهای مختلف ورودی

Number	Structure	Hidden layer	Training			Testing		
			R ²	RMSE (mg/l)	MAE (mg/l)	R ²	RMSE (mg/l)	MAE (mg/l)
1	1-5-1	Mexican hat	0.877	0.844	0.756	0.891	0.784	0.692
2	2-4-1	Mexican hat	0.914	0.765	0.621	0.923	0.736	0.607
3	3-6-1	Mexican hat	0.938	0.712	0.564	0.948	0.701	0.563
4	4-8-1	Mexican hat	0.952	0.681	0.506	0.96	0.668	0.519

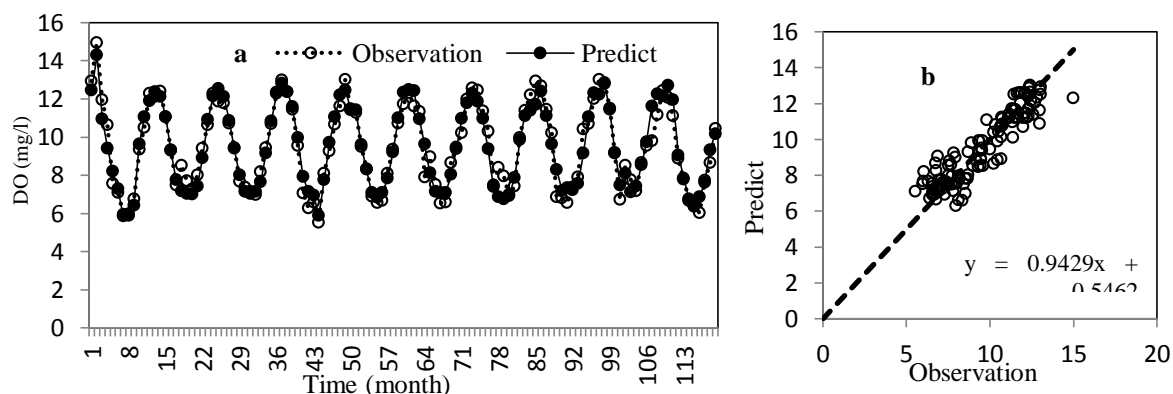


Fig. 3- Observed and predicted values of the wavelet neural network model in training and testing period
 شکل ۳- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل شبکه عصبی موجک در مرحله آموزش و صحت‌سنجی

۲-۳- نتایج ماشین بردار پشتیبان

از سوی دیگر به منظور تخمین میزان اکسیژن محلول در آب رودخانه توسط مدل SVM توابع مختلف کرنل شامل کرنل خطی، چندجمله‌ای و توابع پایه شعاعی که از انواع رایج مورد استفاده در هیدرولوژی است را می‌توان مورد بررسی قرارداد و بهینه‌ترین را انتخاب نمود. نتایج حاصل بررسی کرنل‌ها در جدول ۴ آورده شد. نتایج حاصل از مدل SVM در جدول ۴ ارائه شده است. با توجه به جدول مذکور ساختار ترکیبی شماره ۴ با تابع کرنل پایه شعاعی با بیشترین میزان ضریب همبستگی $R=0.933$ و کمترین جذر مربعات خطا $RMSE=0.821$ mg/l و کمترین میانگین قدر مطلق خطا $MAE=0.645$ mg/l در مرحله صحت‌سنجی جواب بهینه‌تری نسبت به سایر ساختارها از خود نشان داده است. در شکل ۴ نمودار نتایج حاصل از بهترین مدل برای داده‌های بخش آموزش و صحت‌سنجی ارائه شده است. همان‌طور که از شکل (b-۴) مشخص است بیشتر مقادیر تخمینی و مشاهداتی به‌غیر از چند نقطه روی خط نیمساز ($y=x$) قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی است. در شکل (a-۴) نیز تغییرات مقادیر محاسباتی و مشاهداتی نسبت به زمان نشان داده شد، به‌طوری‌که این مدل در تخمین مقادیر کمینه توانایی قابل‌قبولی ندارد. نتایج به‌دست آمده با پژوهش‌های Ghorbani et al. (2016) همخوانی دارد.

۳-۳- نتایج برنامه‌ریزی بیان ژن

استفاده از برنامه‌ریزی بیان ژن به دلیل قابلیت و توانایی مدل در انتخاب متغیرهای مؤثر و حذف متغیرهایی با تأثیر کمتر و همچنین توانایی ارائه رابطه صریح جهت مدل‌سازی میزان اکسیژن محلول در آب رودخانه کامبرلند مدنظر قرار گرفت. بدین منظور هر چهار ورودی برای تعیین متغیرهای معنی‌دار استفاده، و جهت بررسی بیشتر علاوه بر مجموعه چهار عملگر اصلی (F2)، حالت‌هایی بر اساس عملگرهای ریاضی پیش‌فرض برنامه (F1) لحاظ گردیده است. نحوه و دلیل انتخاب این نوع عملگرها با توجه به مطالعات (Khatibi et al., 2012; Ghorbani et al., 2012) صورت گرفته است:

$$F1: \{+, -, *, /, \sqrt{\quad}, \text{Exp}, \text{Ln}, 2^{\quad}, 3^{\quad}, \sqrt[3]{\quad}, \text{Sin}, \text{Cos}, \text{Atan}\} \quad (20)$$

$$F2: \{+, -, *, /\} \quad (21)$$

نتایج به‌دست آمده از اجرای مدل برنامه‌ریزی بیان ژن با هر دو عملگر و ساختارهای مختلف در جدول ۵ ارائه شده است. مشاهده می‌گردد که ساختار ترکیبی شماره ۴ و عملگر F2 در هر دو مرحله آموزش و صحت‌سنجی با بیشترین ضریب همبستگی $R=0.911$ و جذر میانگین مربعات خطا $RMSE=0.906$ mg/l و میانگین قدر مطلق خطا $MAE=0.694$ نسبت به سایر ساختارها و عملگرها از دقت بیشتری برخوردار بوده است.

Table 4- The analysis of support vector machine model results for different input structures

جدول ۴- تحلیل نتایج مدل ماشین بردار پشتیبان برای ساختارهای مختلف ورودی

Number	Kernel	Training			Testing		
		R ²	RMSE (mg/l)	MAE (mg/l)	R ²	RMSE (mg/l)	MAE (mg/l)
1	RBF	0.735	1.502	1.141	0.718	1.78	1.268
	Poly	0.734	1.503	1.143	0.716	1.782	1.281
	lin	0.628	1.903	1.601	0.715	1.786	1.36
2	RBF	0.91	0.916	0.695	0.907	0.936	0.733
	poly	0.908	0.92	0.698	0.874	1.141	0.907
	lin	0.889	1.014	0.779	0.9	1.017	0.845
3	RBF	0.921	0.864	0.639	0.908	1.008	0.827
	poly	0.92	0.868	0.642	0.905	1.009	0.828
	lin	0.895	0.988	0.738	0.9	1.015	0.834
4	RBF	0.948	0.703	0.527	0.933	0.821	0.645
	poly	0.936	0.776	0.574	0.924	0.863	0.701
	lin	0.936	0.776	0.574	0.923	0.863	0.701

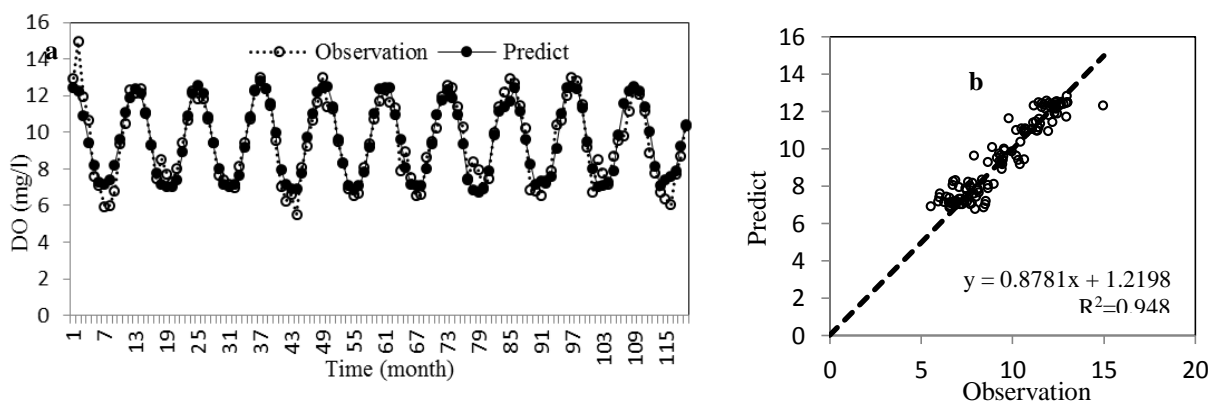


Fig. 4- Observed and predicted values of the support vector machine model in training and testing
شکل ۴- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل ماشین بردار پشتیبان در مرحله آموزش و صحت‌سنجی

۴-۳- مقایسه عملکرد مدل‌ها

همان‌طور که در جدول ۶ مشاهده می‌شود با انتخاب جواب بهینه هر کدام از مدل‌ها و مقایسه آن‌ها با یکدیگر مشخص شد که هر سه مدل با دقت خوبی و خطای کمتر می‌توانند میزان اکسیژن محلول در آب رودخانه کامبرلند را شبیه‌سازی کنند. که از بین مدل‌های به‌کاررفته مدل شبکه عصبی موجک، بیشترین دقت $R=0.960$ و کمترین ریشه میانگین مربعات خطا $RMSE=0.668$ mg/l و کمترین میانگین قدر مطلق خطا $MAE=0.519$ mg/l در مرحله صحت‌سنجی را از خود نشان داده است. در شکل ۵ نتایج هر سه مدل نسبت به مقادیر مشاهداتی در طی زمان نشان داده شد. از سوی دیگر طبق جدول ۶ نتایج نشان داد در هر سه مدل، ساختارهای ترکیبی نسبت به سایر ساختارها عملکرد قابل‌قبولی دارند، به‌گونه‌ای که با افزایش دمای آب رودخانه منجر به کاهش اکسیژن محلول در آب می‌گردد. همچنین کاهش نور، سبب کاهش مقدار اکسیژن آزاد شده توسط گیاهان می‌شود بطوریکه اگر نور به شکل کامل به گیاهان کف نرسد، تولید اکسیژن

بنابراین، برنامه‌ریزی بیان ژن با عملگر F2 شامل چهار عملگر اصلی ریاضی، باوجود رابطه ساده ریاضی بیشترین دقت را در مدل‌سازی اکسیژن محلول در آب رودخانه کامبرلند داشته است. نمودار مقادیر بهینه محاسباتی و مشاهداتی برنامه‌ریزی بیان ژن مربوط به مرحله آموزش و صحت‌سنجی در شکل ۵ نشان داده شده است. همان‌طور که از شکل مشخص است مدل برنامه‌ریزی ژن نتوانسته است در تخمین نقاط بیشینه و کمینه دقت قابل‌قبولی ارائه دهد. در تبیین این نتایج می‌توان بیان نمود که در این پژوهش جهت افزایش دقت مدل عملگرهای \sin ، \cos و غیره کاربرد ندارد و نیز با توجه به دقت و سادگی، مدل حاصل از چهار عمل اصلی ریاضی $\{+, -, *, /\}$ جهت تخمین میزان اکسیژن محلول در آب رودخانه کامبرلند پیشنهاد شد (Ghorbani et al., 2012).

Table 5- The analysis of gene expression programming model results for different input structures

جدول ۵- تحلیل نتایج مدل برنامه‌ریزی بیان ژن برای ساختارهای مختلف ورودی

Number	operator	Training			Testing		
		R ²	RMSE (mg/l)	MAE (mg/l)	R ²	RMSE (mg/l)	MAE (mg/l)
1	F1	0.715	1.552	1.210	0.718	1.682	1.331
	F2	0.710	1.562	1.203	0.752	1.613	1.237
2	F1	0.714	1.553	1.220	0.737	1.676	1.318
	F2	0.757	1.450	1.128	0.684	1.780	1.344
3	F1	0.870	1.104	0.854	0.913	1.062	0.879
	F2	0.895	0.992	0.740	0.923	0.936	0.781
4	F1	0.875	1.077	0.864	0.909	1.165	0.893
	F2	0.915	0.896	0.709	0.911	0.906	0.694

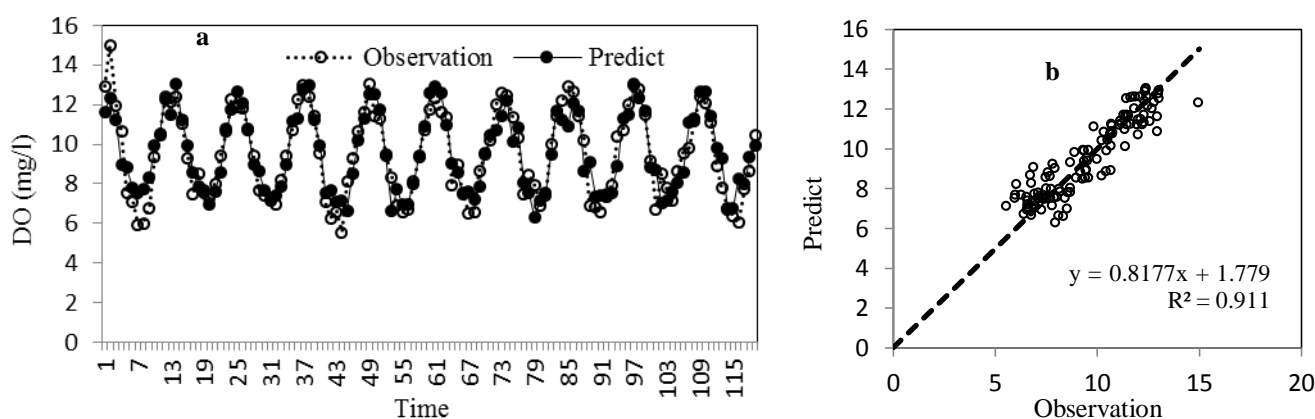


Fig. 5- Observed and predicted values of the gene expression programming model in training and testing period

شکل ۵- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل برنامه‌ریزی بیان ژن در مرحله آموزش و صحت‌سنجی

بیشتر می‌شود، دقت کار افزایش می‌یابد. بسامدهای مراحل پایین‌تر تجزیه، نویز بیشتری دارند، ولی با افزایش سطح تجزیه از میزان نویز کاسته شده و سیگنال نرم‌تر می‌شود. همچنین در شکل ۷ نمودار ویولن هر سه مدل نسبت به مقادیر مشاهداتی قابل مشاهده است که در این شکل میانگین مقادیر محاسباتی مدل برنامه‌ریزی بیان ژن کمتر از سایر مدل‌ها می‌باشد و نیز با توجه به کران پایین مدل شبکه عصبی موجک عملکرد بهتری نسبت به سایر مدل‌ها در تخمین مقادیر مینیمم از خود نشان داده است. در نهایت اختلاف مقادیر مشاهداتی میزان اکسیژن محلول در آب و محاسباتی بهینه مدل‌ها به صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به نمونه داده‌های ثبت شده دوره آماری ترسیم گردید (شکل ۸). همان‌طور که در این شکل دیده می‌شود برای هر سه مدل بیشتر خطاها در باند $\pm 10\%$ قرار گرفته است. که در میان مدل‌های مذکور (شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن)، مدل شبکه عصبی موجک کمترین میزان خطا را داشته است.

توسط این گیاهان متوقف شده و باکتری‌های موجود، اکسیژن را مصرف می‌کنند. از طرفی دیگر تأثیر دبی رودخانه بر روی میزان اکسیژن محلول در آب سبب می‌گردد که با افزایش دبی رودخانه غلظت آب ناشی از ورود پساب افزایش نیافته و اکسیژن محلول در آب کاهش نیابد و همان‌طور که در شکل ۶ قابل مشاهده است هر سه مدل در تخمین بیشتر مقادیر عملکرد قابل قبولی دارند. گرچه مدل شبکه عصبی موجک در پیش‌بینی مقادیر حداقل، حداکثر و میانی میزان اکسیژن محلول در آب رودخانه، عملکرد قابل قبولی از خود نشان داده است. این نتایج با پژوهش‌های (Alizadeh et al. (2015) و (Radmanesh et al. (2016) نیز همخوانی دارد. در تبیین این نتایج می‌توان بیان نمود که تبدیل موجکی با جدا نمودن سیگنال به فرکانس‌های بالا و پائین، ویژگی‌های چند مقیاسی سیگنال را در اختیار داشته و دقت مدل را تا حد قابل توجهی، بالا می‌برد. سیگنال‌های بالا گذر و پایین گذر حاصل از تجزیه موجک، برآزش بسیار خوبی با معادلات مجموع سینوسی دارند که هر چه تعداد مراتب این معادلات

Table 6- The statistical analysis of wavelet neural network, support vector machine and gene expression programming results for different input structures

جدول ۶- تحلیل آماری نتایج مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن برای ساختارهای مختلف ورودی

Models	Training			Testing		
	R ²	RMSE (mg/l)	MAE (mg/l)	R ²	RMSE (mg/l)	MAE (mg/l)
WNN	0.952	0.681	0.506	0.96	0.668	0.519
SVM	0.948	0.703	0.527	0.933	0.821	0.645
GEP	0.915	0.896	0.709	0.911	0.906	0.694

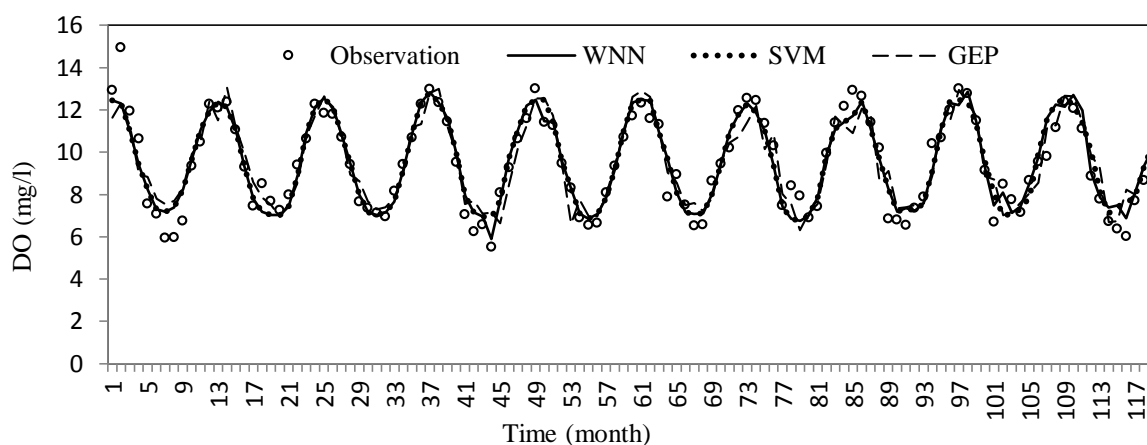


Fig. 6- Observed and predicted values of the WNN, SVM and GEP models in training and testing period
 شکل ۶- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در مرحله آموزش و صحت‌سنجی

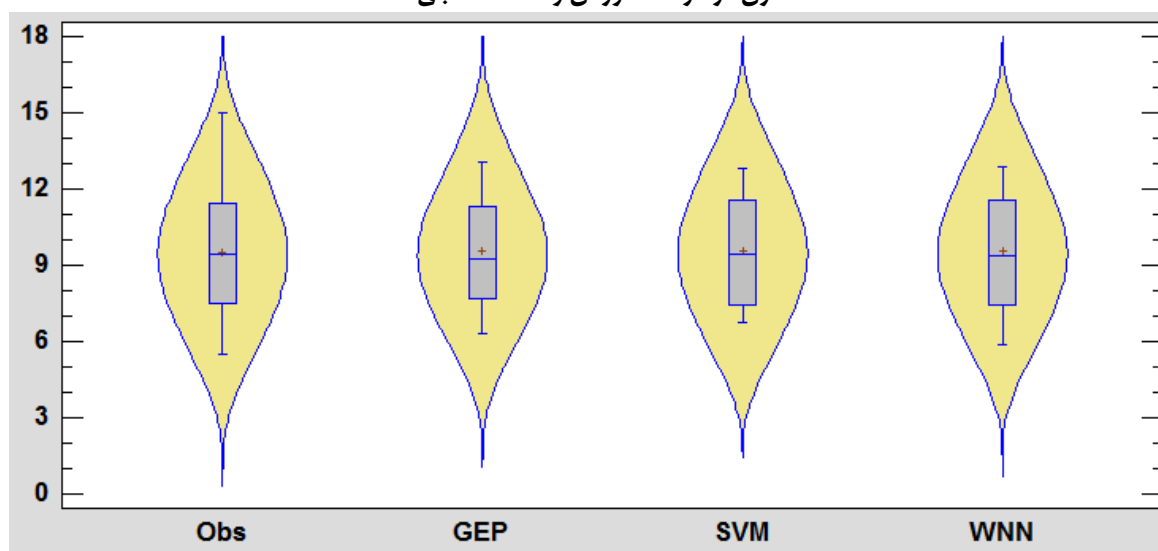


Fig. 7- Violin plot observed and predicted values of the WNN, SVM and GEP models in training and testing period
 شکل ۷- نمودار ویولن مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در مرحله آموزش و صحت‌سنجی

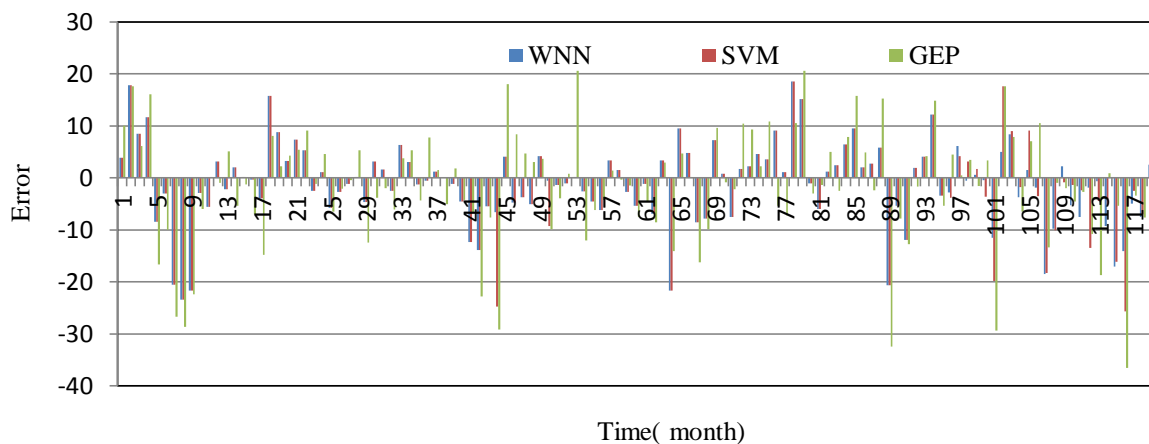


Fig. 8- The graph of optimal error for three models as a percentage of mean observed values
شکل ۸ - نمودار خطای بهینه هر سه مدل به صورت درصدی از میانگین مقادیر مشاهداتی

موجک جهت پیش‌بینی پارامتر اکسیژن محلول در آب، می‌تواند به‌عنوان ابزاری مناسب در اتخاذ تصمیمات مدیریتی مورد استفاده قرار گیرد. علاوه بر آن تصمیم در خصوص لزوم استفاده از ابزارهای هوادهی برای افزایش اکسیژن محلول در آب به مقادیر برآوردی توسط مدل وابسته است. همچنین پیشنهاد می‌گردد که استفاده از مدل‌های هیبریدی (حالت ترکیب مدل‌های مختلف با الگوریتم‌های بهینه‌سازی جدید) نیز جهت پیش‌بینی پارامتر اکسیژن محلول در آب مورد بررسی و ارزیابی قرار گرفته و در نهایت با روش‌های موجود مقایسه شود.

پی‌نوشت‌ها

1. Wavelet Neural Network
2. Gene Expression Programming
3. Support Vector Machine
4. Hilo
5. Kodaikanal
6. Eymir
7. Langat
8. Kentucky
9. Crossover
10. Mutation

۵- مراجع

- Alizadeh MJ, Kavianpour MR (2015) Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin* 98(1-2):171-8
- Banejad H, Kamali M, Amir Moradi K, Olyaei A (2013) Forecasting some of the qualitative parameters of rivers using Wavelet Artificial Neural Network Hybrid (W-ANN) model (case of study: Jajroud river of Tehran and Gharaso river of Kermanshah). *Iranian*

۴- نتیجه‌گیری

در این تحقیق سعی گردید عملکرد سه مدل شبکه عصبی موجک، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان جهت شبیه‌سازی میزان اکسیژن محلول در آب رودخانه کامبرلند واقع در ایالت تنسی، با استفاده از داده‌های ماهانه مورد ارزیابی قرار گیرد. مقادیر جریان پیش‌بینی شده در مدل‌های مذکور، با استفاده از معیارهای ارزیابی و بر اساس مقادیر جریان مشاهداتی مورد مقایسه قرار گرفت. نتایج حاصل از این تحقیق نشان داد که در هر سه مدل مورد بررسی شامل شبکه عصبی موجک، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان، ساختارهای ترکیبی (ترکیب شماره ۴) نتایج بهتری نسبت به سایر ساختارها ارائه می‌دهد، که نشان‌دهنده انتخاب مناسب پارامترهای ورودی و تأثیر تمامی پارامترهای اندازه‌گیری شده در مدل‌سازی اکسیژن محلول در آب می‌باشد. همچنین با توجه به معیارهای ارزیابی مشاهده گردید که هر سه مدل مورد بررسی، می‌توانند با دقت نسبتاً بالایی میزان اکسیژن محلول در آب رودخانه را پیش‌بینی نمایند. در این میان، مدل شبکه عصبی موجک دقت بالاتر و خطای کمتری نسبت به مدل‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن به‌خصوص در پیش‌بینی مقادیر حداقل و حداکثر از خود نشان داده است، که دقت بالای شبکه عصبی موجک ناشی از جدا نمودن سیگنال به فرکانس‌های بالا و پایین بوده است. در مجموع این تحقیق نشان داد که استفاده از مدل شبکه عصبی موجک می‌تواند در زمینه تخمین میزان اکسیژن محلول در آب رودخانه‌ها مؤثر باشد. با توجه به اینکه تصمیم‌گیری برای تناسب کیفیت آب‌های سطحی و پیاده‌سازی استراتژی‌های مدیریتی در بسیاری مصارف (به‌ویژه بخش شیلات و آبی‌پروری) قویاً به میزان اکسیژن محلول در آب وابسته است لذا استفاده از روش شبکه عصبی

- Hamel L (2009) Knowledge discovery with support vector machines. Hoboken, N.J. John Wiley
- Khatibi R, Naghipour L, Ghorbani MA, Aalami MT (2012) Predictability of relative humidity by two artificial intelligence techniques using noisy data from two Californian gauging stations. *Neural Computing and Application* 23(7):643-941
- Kisi O, Karahan M, Sen Z (2006) River suspended sediment modeling using fuzzy logic approach. *Hydrol Process* 20:4351-4362
- Lin JY, Cheng CT, Chau KW (2006) Using support vector machines for long-term discharge prediction. *Hydrolog Sci J* 51(3):599–612
- Liong SY, Sivapragasam C (2002) Flood stage forecasting with support vector machines. *J Am Water Resour* 38(4):173–186
- Lopes JF, Dias JM, Cardoso AC, Silva CIV (2005) The water quality of the Ria de Aveiro lagoon, Portugal: from the observations to the implementation of a numerical model. *Mar Environ Res* 60:594-628
- Nagy H, Watanabe K, Hirano M (2002) Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulics Engineering* 128:558-559
- Nourani V, Kisi Ö, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. *Journal of Hydrology* 402(3): 41–59
- Nozari H, Tavakoli F (2017) Evaluation of the efficiency of conventional and computerized methods for reconstruction of monthly flow time series in hydrometric stations. *Iran-Water Resources Research* 13(4):174-178 (In Persian)
- Radmanesh F, Pourhaghi A, Solgi A (2016) Improvement of artificial neural network modeling using wavelet transformation and PCA method for modeling and predicting biosphere oxygen (BOD). *Journal of Ecohydrology* 3(4):569-585 (In Persian)
- Radwan M, Willems P, El-Sadek A, Berlamont J (2003) Modelling of dissolved oxygen and biochemical oxygen demand in river water using a detailed and simplified model. *Int J River Basin Manage* 1(2):97-103
- Raheli B, Alami MT, El-Shafie A, Deo R (2017) Uncertainty assessment of the Multilayer Perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. *Environmental Earth Sciences* 76(503):3-16
- journal of Health and Environment 6(3):277-294 (In Persian)
- Chapman D (1992) *Water Quality Assessments*. Chapman and Hall Ltd, London, UK
- Chu HB, Lu WX, Zhang L (2013) Application of artificial neural network in environmental water quality assessment. *Journal of Agricultural Science and Technology* 15(2):343-356
- Dehghani R, Ghorbani MA, Teshnehlab M, Rikhtegar A, Asadi A (2015) Comparison and evaluation of Bayesian neural network, gene gramming, support vector machine and multiple expression prolinear regression in river discharge estimation (case study: Sufi Chay Basin). *Iranian journal of Irrigation and Water* 20(5):66-85 (In Persian)
- Dogan E, Lent Sengorur B, Koklu R (2009) Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management* 90:1219-35
- Eskandari A, Nouri R, Meragi H, Kiaghaderi A (2012) Developing an appropriate model based on artificial neural network and support vector machine for prediction of timely 5 day biochemical oxygen demand. *Journal of Ecology* 38(1):71-82 (In Persian)
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems* 13(2):87–129
- Ghorbani MA, Dehghani R (2017) Comparison of Bayesian neural network, artificial neural network gene expression programming in river water quality (case study: Belkhviachay river). *Watershed Management Research* 8(15):13-24 (In Persian)
- Ghorbani MA, Ahmadzadeh H, Isazadeh M, Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences* 75(476):3-14
- Ghorbani MA, Azani A, Mahmoudi S (2015) Rainfall-Runoff modeling using hybrid intelligent models. *Iran-Water Resources Research* 11(2):146-150 (In Persian)
- Ghorbani MA, Khatibi R, Asadi H, Yousefi P (2012) Inter-comparison of an evolutionary programming model of suspended sediment time-series with other local model. *INTECH*.doi. org/10.5772/47801: 255-282
- Ghorbani MA, Salehi A (2011) Use of gene expression planning to study the variation of groundwater quality data with fluctuations in the water level in Isfahan plain. *Sixth National Congress on Civil Engineering* (In Persian)

- Vapnik VN (1995) *The Nature of Statistical Learning Theory*. Springer, New York
- Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
- Wang D, Safavi AA, Romagnoli JA (2000) Wavelet-based adaptive robust M-estimator for non-linear system identification. *AIChE Journal* 46(4):1607-1615
- Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396(4):128–138
- Shin S, Kyung D, Lee S, Taik & Kim J, Hyun J (2005) An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28(4):127-135
- Solgi A, Pourhaghi A, Zarei H, Ansari H (2017) Modeling and forecast biological oxygen demand (BOD) using combination support vector machine with wavelet transform. *Journal of Water and Soil* 31(1):86-100 (In Persian)
- Vapnik V, Chervonenkis A (1991) The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognition and Image Analysis* 1(3):283-305