



Comparison of artificial neural network methods and support vector machine in predicting water quality parameters of Dinachal river, IRAN

Sina Asadpour Lomer¹ , Kumars Ebrahimi² 

1. Department of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran, Email: Sina.Asadpour@ut.ac.ir
2. Corresponding Author, Department of Renewable Energies and Sustainable Resources Engineering, University of Tehran, Tehran, Iran. Email: EbrahimiK@ut.ac.ir

Article Info	ABSTRACT
<p>Article type: Research Article</p> <p>Article history:</p> <p>Received: Jan. 8, 2024</p> <p>Revised: Feb. 5, 2024</p> <p>Accepted: Feb. 5, 2024</p> <p>Published online: Sep. 2024</p> <p>Keywords: ANN, Dinachal River, SVR, Water Quality Prediction.</p>	<p>Predicting water quality parameters plays a crucial role for monitoring of ecosystems of rivers and their sustainability. Alongside this, conventional prediction models are not able to capture the non-linearity and non-stationary inherence of water quality datasets. In recent years, the rapid development of machine learning methods has transformed the water quality prediction fields. In this study, water quality parameters for the Dinachal River in Guilan province have been assessed and predicted. Two models based on Artificial Neural Networks (ANN) and Support Vector Regression (SVR) were utilized to predict nine water quality parameters as TDS, EC, pH, Cl, SO₄, HCO₃, Ca, Mg, and Na with monthly timesteps between 2006 and 2018. Then, the model's performance was evaluated using RMSE, MSE, and MAPE indices. According to the results, the SVR model was superior in predicting TDS and Mg parameters with an RMSE Index of 2.03 and 0.062, respectively. Simultaneously, ANN had a slightly better accuracy in the prediction of remaining parameters. However, prediction results for both models in the case study were satisfactory. In addition, the SVR model predicted TDS with a MAPE of 0.007, which was the best compared to other parameters. At the same time, the ANN model had better performance in predicting EC with a MAPE of 0.001. Prediction results for Cl had also the lowest accuracy among water quality parameters with an RMSE of 0.055 and 0.052 for SVR and ANN, respectively. Methods utilized in this study can be effective in predicting water quality parameters of Dinachal river.</p>

Cite this article: Asadpour Lomer, S., & Ebrahimi, K. (2024) Comparison of Artificial Neural Network Methods and Support Vector Machine in Predicting Water Quality Parameters of Dinachal River, IRAN, *Iranian Journal of Soil and Water Research*, 55 (7), 1033-1046.

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2024.370818.669644>



قیاس روش‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان در پیش‌بینی پارامترهای کیفیت آب رودخانه دیناچال

سینا اسدیپور لمر^۱، کیومرث ابراهیمی^۲

۱. گروه مهندسی آبیاری و آبادانی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: Sina.Asadpour@ut.ac.ir
۲. نویسنده مسئول، گروه مهندسی انرژی‌های نو و منابع پایدار، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران.
رایانامه: EbrahimiK@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۱۰/۱۸

تاریخ بازنگری: ۱۴۰۲/۱۱/۱۶

تاریخ پذیرش: ۱۴۰۲/۱۱/۱۶

تاریخ انتشار: مهر ۱۴۰۳

واژه‌های کلیدی:

پیش‌بینی کیفیت آب،

شبکه عصبی مصنوعی،

ماشین بردار پشتیبان،

رود دیناچال.

پیش‌بینی پارامترهای کیفیت آب نقش بسیار مهمی در پایش اکوسیستم رودخانه‌ها و پایداری آن ایفا می‌کند. از طرفی، مدل‌های پیش‌بینی سنتی به خوبی ذات غیرخطی و غیرایستای متغیرهای کیفیت آب را نشان نمی‌دهند. در سال‌های اخیر، توسعه سریع شبکه‌های عصبی مصنوعی بحث‌های مرتبط با پیش‌بینی کیفیت آب را متحول نموده است. در این مطالعه، پارامترهای کیفی رودخانه دیناچال به‌عنوان یکی از رودخانه‌های حیاتی در استان گیلان مورد ارزیابی و پیش‌بینی قرار گرفته است. دو مدل مبتنی بر شبکه عصبی مصنوعی و ماشین بردار پشتیبان رگرسیونی، نه پارامتر کیفی کل‌مواد جامد محلول، هدایت الکتریکی، اسیدیته، SO_4 ، Cl ، Ca ، Mg و Na با گام‌های زمانی یک ماه طی سال‌های ۱۳۸۵ تا ۱۳۹۷ پیش‌بینی شد و کارایی مدل‌ها توسط آماره‌های ارزیابی $RMSE$ ، MSE ، $MAPE$ بررسی و مقایسه شد. طبق نتایج به‌دست آمده، مدل SVR به‌ترتیب با $(RMSE=2/0.3)$ و $(RMSE=0/0.62)$ کارایی بهتری در پیش‌بینی کل‌مواد جامد محلول و غلظت منیزیم نسبت به مدل شبکه عصبی مصنوعی نشان داد. از طرف دیگر، مدل شبکه عصبی مصنوعی در پیش‌بینی پارامترهای دیگر نسبتاً موفق‌تر بود. با این حال، کارایی هر دو مدل در پیش‌بینی پارامترهای کیفی رودخانه دیناچال مناسب ارزیابی شد. همچنین مدل SVR با ضریب $MAPE$ برابر با $0/0.07$ و مدل شبکه عصبی مصنوعی با ضریب $MAPE$ برابر با $0/0.01$ در پیش‌بینی پارامترهای کل‌مواد جامد محلول و هدایت الکتریکی بهترین کارایی را داشتند. در مقابل، هر دو مدل SVR و ANN در پیش‌بینی پارامتر کلر ضعیف‌ترین کارایی را با وجود $RMSE$ برابر با $0/0.55$ و $0/0.52$ از خود نشان دادند. روش‌های به‌کار گرفته شده در این مطالعه می‌توانند در پیش‌بینی کیفیت آب دیناچال مؤثر واقع شود.

استاد: اسدیپور لمر؛ سینا، ابراهیمی؛ کیومرث، (۱۴۰۳) قیاس روش‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان در پیش‌بینی پارامترهای کیفیت آب رودخانه دیناچال،

مجله تحقیقات آب و خاک ایران، ۵۵ (۷) ۱۰۳۳-۱۰۴۶.

ناشر: مؤسسه انتشارات دانشگاه تهران.

© نویسندگان.



DOI: <https://doi.org/10.22059/ijswr.2024.370818.669644>

مقدمه

با توجه به افزایش روزافزون جمعیت و گسترش شهرها و روستاها و در نتیجه افزایش نیاز مصرف و برداشت از آب سطحی و زیرزمینی، توجه به حفاظت و صیانت از این منابع ارزشمند بیش از گذشته اهمیت پیدا کرده است (ستاری و همکاران، ۱۳۹۲). رودخانه‌ها همواره از مهم‌ترین منابع تأمین آب شرب و کشاورزی هستند که نقش به‌سزایی در تأمین سلامت انسان و محیط‌زیست دارند؛ اما امروزه با تخلیه انواع مختلف آلاینده‌های کشاورزی، صنعتی و پساب‌های شهری به رودخانه‌ها موجب شده که این منابع مهم، گاهی اوقات به‌عنوان یکی از مراکز بحرانی آلودگی‌ها مطرح باشند (شاهی نژاد و همکاران، ۱۴۰۰). به‌طور معمول آب‌های سطحی بیش‌تر از بقیه آب‌ها در برابر آلودگی‌های محیطی آسیب‌پذیر هستند؛ بنابراین به‌دلیل افزایش معنادار منابع آلودگی، پایش مستمر و پیش‌بینی وضعیت کیفیت این منابع از اهمیت به‌سزایی برخوردار است (ویسی و همکاران، ۱۳۹۶). به‌منظور ارزیابی کیفیت آب‌های سطحی در دوره‌های زمانی گوناگون، از این منابع نمونه‌برداری می‌شود و سپس در آزمایشگاه‌های سنجش کیفیت آب، میزان پارامترهای کیفی مختلف در آن‌ها اندازه‌گیری و مورد ارزیابی قرار می‌گیرند. به‌عبارتی‌دیگر سنجش و ارزیابی پارامترهای کیفی حوضه‌های آبریز امری ضروری است (Huang et al., 2018). از طرفی، مدل‌های پیش‌بینی سنتی به‌خوبی ذات غیرخطی و غیر ایستای متغیرهای کیفیت آب را نشان نمی‌دهند. در این خصوص تمرکز بر روش‌های مختلف پایش و پیش‌بینی کیفی رودخانه‌ها که مبتنی بر الگوهای گوناگون ریاضی هستند و با کمک یادگیری ماشین‌ها توسعه داده‌شده است (Mahapatra et al., 2011).

در ادامه، سوابق مطالعاتی داخلی و خارجی مرتبط با تحقیق حاضر تشریح می‌شود و پس از معرفی منطقه مطالعاتی، مواد و روش‌های به‌کاررفته در این پژوهش بیان می‌شود. در انتها نیز نتایج به‌همراه نتیجه‌گیری نهایی این پژوهش ارائه می‌شود.

وضعیت خشک‌سالی ایستگاه سینوپتیک شهر سقز را به‌عنوان یکی از مهم‌ترین ایستگاه‌های جنوبی حوضه آبریز دریاچه ارومیه، با استفاده از شاخص بارش - تبخیر و تعرق استاندارد شده^۱ (SPEI) و مدل ماشین بردار پشتیبان (SVM) بررسی شد. نتایج نشان داد در هر دو روش مذکور، با استفاده از پنج گام زمانی خروجی‌های مدل دقیق‌تر و بهینه‌تر است. همچنین دقت پیش‌بینی هر دو مدل رابطه مستقیمی با افزایش مقیاس شاخص تبخیر و تعرق استاندارد شده دارد (عباسی و همکاران، ۱۳۹۷).

برای پیش‌بینی فصلی پارامترهای هدایت الکتریکی و یون کلر، از مدل خود هم‌بسته‌متحرک فصلی انباشته^۲ (SARIMA) استفاده شد. نتایج نشان داد که پارامترهای هدایت الکتریکی و غلظت یون کلر، دارای تناوب فصلی ۱۲ ماهه و فاقد مؤلفه روند هستند. همچنین پیش‌بینی‌های این مدل از لحاظ آماری با مقادیر ثبت‌شده، تفاوت معناداری ندارد (جمالی و ابراهیمی، ۱۳۹۰).

با استفاده از پارامترهای نسبت جذب سدیم و یون کلر، روش‌های ماشین بردار پشتیبان و شبکه عصبی را یکدیگر مقایسه شدند. نتایج نشان داد که مدل بردار پشتیبان نسبت به روش شبکه عصبی ضریب همبستگی بالاتری دارد و همچنین ریشه میانگین مربعات خطا کمتری دارد و در نهایت مدل بردار پشتیبان برای پیش‌بینی کیفیت و مقادیر پارامترهای کیفی رودخانه سفیدرود پیشنهاد می‌شود (حسین‌پناهی و همکاران، ۱۴۰۰).

به‌منظور پیش‌بینی غلظت اکسیژن محلول رودخانه ساواناه امریکا، از روش‌های یادگیری عمیق و یادگیری ماشین استفاده شد و نتایج حاصل از هر دو روش با یکدیگر قیاس شد. نتایج نشان داد با حذف پارامتر دمای آب؛ سنجش‌های ارزیابی به میزان ۱۴ درصد کاهش و مقدار خطا ۱۰۰ درصد افزایش داشته است. بنابراین دمای آب به‌عنوان تأثیرگذارترین پارامتر در پیش‌بینی اکسیژن محلول در آب معرفی شد (روشنگر و داودی، ۱۴۰۱).

جریان رودخانه باراندوزچای با روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان پیش‌بینی شد. نتایج نشان داد که میزان دقت پیش‌بینی در هر دو روش تفاوت معناداری نداشته و همچنین میزان دقت در هر دو روش رابطه غیرمستقیمی با افزایش مقدار جریان عبوری رودخانه دارد. به این مفهوم که با افزایش مقدار جریان عبوری رودخانه، میزان دقت در هر دو روش کاهش می‌یابد. به دلیل عدم تفاوت معنادار در میزان دقت در هر دو مدل، روش بردار پشتیبان به دلیل سهولت در اجرا نسبت به روش برنامه‌ریزی ژنتیک، برای استفاده پیشنهاد می‌شود (احمدی و همکاران، ۱۳۹۳).

غلظت پتاسیم موجود در شورابه‌های پلایای شهرستان خور بیابانک اصفهان، با استفاده از روش بردار پشتیبان و جنگل تصادفی پیش‌بینی شد و میزان تأثیر پارامترهای مختلف بر یکدیگر، مورد بررسی قرار گرفت. نتایج نشان داد ضریب تعیین برای هر دو مدل به ترتیب

۰/۹۹ و ۰/۹۷ بوده که بیانگر دقت بالا در هر دو روش است (ایرجی و همکاران، ۱۴۰۲).

با ترکیب مدل‌های ماشین بردار پشتیبان و الگوریتم‌های بهینه‌سازی، پارامترهای کیفیت آب رودخانه کلانگ مالزی مطالعه شد. در این پژوهش برای نخستین بار شاخص کیفیت آب (WQI) با ترکیب مدل‌های مذکور بهینه‌سازی شد که نتایج نشان داد، مدل بردار پشتیبان نسبت به سایر روش‌های مورد استفاده در این تحقیق، ریشه میانگین مربعات خطای کمتر و سهولت اجرای بیشتری دارد (SeeLeng et al., 2018).

با استفاده از روش بردار پشتیبان و با بهره‌گیری از پانزده پارامتر کیفیت آب رودخانه وادی صف واقع در کشور الجزایر، میزان دقت پیش‌بینی پارامترهای مختلف و عملکرد مدل ارزیابی شد. طبق خروجی‌های مدل با اعداد (همبستگی در مرحله آموزش ۰/۸۲) و (ریشه میانگین مربعات خطا در مرحله آموزش ۵/۱۷) و (همبستگی در مرحله آزمون ۰/۸۱) و (ریشه میانگین مربعات خطا در مرحله آزمون ۵/۵۵)، می‌توان به پیش‌بینی بهینه مدل در هر دو مرحله آموزش و آزمون اذعان داشت. همچنین با کاهش تعداد متغیرهای ورودی، دقت مدل افزایش می‌یابد (Sakka et al., 2022).

از مدل بردار پشتیبان، برای پیش‌بینی کیفیت آب رودخانه لگات در کشور مالزی استفاده شد. از طریق تجزیه و تحلیل شش پارامتر کیفی، تحت دو سناریوی متفاوت، مقادیر اکسیژن محلول (DO) را پیش‌بینی شد. نتایج نشان داد ضریب همبستگی بین پارامترهای پیش‌بینی شده ۰/۹۱ تا ۰/۹۴ است و میزان خطا در این مدل‌سازی برابر با ۰/۰۱۶ است (Yahya et al., 2019).

برای پایش و حفاظت از کیفیت آب رودخانه و مدل‌سازی کیفیت آب، از مدل ماشین بردار پشتیبان (SVM) و شبکه عصبی مصنوعی (ANN)، جهت پیش‌بینی کل مواد جامد و کل جامدات معلق استفاده شد. نتایج نشان داد که میزان ریشه میانگین مربعات خطا در روش بردار پشتیبان نسبت به روش شبکه عصبی مصنوعی کمتر و میزان همبستگی خروجی برای پارامتر کل مواد جامد معلق در روش شبکه عصبی مصنوعی بیشتر از روش ماشین بردار پشتیبان است (Rizal et al., 2022).

با به‌کارگیری یک رویکرد مبتنی بر ماشین بردار پشتیبان با ادغام الگوریتم شبکه عصبی جهت طبقه‌بندی کیفی رودخانه چائو فرامیا با استفاده از پارامترهای کیفیت آب از جمله نیترات و شوری و کل مواد جامد، توابع ریاضی بهینه و مناسب این امر بررسی شد. نتایج نشان داد، با افزایش تعداد توابع ریاضی به‌کاررفته در مدل‌سازی، میزان دقت طبقه‌بندی کیفیت آب نیز افزایش می‌یابد. همچنین کم‌ترین میزان دقت شبیه‌سازی مربوط به پارامتر نیترات عنوان شد (Sillberg et al., 2021).

به‌منظور پیش‌بینی پارامترهای کیفیت آب رودخانه داهیتینگ، از روش شبکه عصبی مصنوعی استفاده شد. با داشتن مقادیر پارامترهای کیفی آب رودخانه، مقادیر کل نیتروژن و کل فسفر پیش‌بینی شد. نتایج نشان داد میزان همبستگی پارامتر کل نیتروژن از مقدار همبستگی پارامتر کل فسفر بیشتر و برابر با ۰/۹۴ است (Yan et al., 2021).

با استفاده از روش شبکه عصبی مصنوعی، امکان پیش‌بینی پارامترهای کیفیت آب و تاثیر پذیری از ویژگی‌های کاربری زمین، با رویکردی تحقیقی بررسی و ارزیابی شد. نتایج نشان داد که میزان خطای پیش‌بینی پارامترهای کیفیت آب با این روش برابر با ۰/۶۷ است (Gorashi et al., 2012).

به‌منظور بررسی و پیش‌بینی کیفیت آب رودخانه هایه و با استفاده از روش شبکه عصبی مصنوعی، پارامترهای کیفیت آب این رودخانه بررسی و شبیه‌سازی شد. در اولین ارزیابی کیفی رودخانه، برخلاف انتظار، مقادیر اکسیژن محلول کمتر و نیتروژن کل بیشتر از تصور بود که دلیل آن وقوع پدیده دنیتریفیکاسیون اعلام شد. نتایج نشان داد، ضریب همبستگی برای پارامتر اکسیژن محلول از بقیه پارامترها بیشتر و برابر با ۰/۷۹ است (Li et al., 2023).

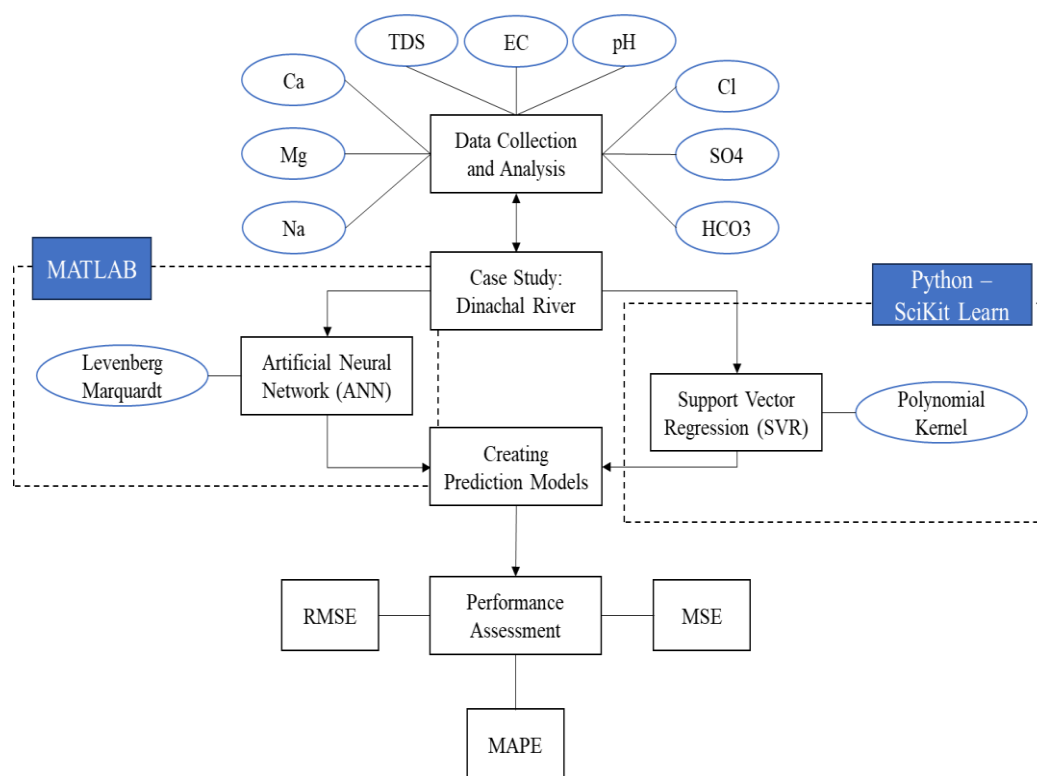
با توجه به وابستگی فعالیت‌های صنعتی، کشاورزی، تأمین آب شرب و زیست‌محیطی شهرستان رضوانشهر و شهر پره‌سر به جریان‌های آبی رودخانه دیناچال، پایش و پیش‌بینی شاخص‌های کیفی این رودخانه از اهمیت ویژه‌ای برای ساکنان این منطقه برخوردار است. در این راستا در تحقیق حاضر از روش‌های پیشرفته مبتنی بر یادگیری ماشین‌ها و به‌منظور تدقیق بیشتر از دو روش شبکه عصبی مصنوعی^۱ (ANN) با کمک کد نویسی متلب و روش ماشین بردار پشتیبان^۲ (SVM) با کمک کدنویسی پایتون استفاده شده است. سپس نتایج با داده‌های مشاهداتی مقایسه و ارزیابی شده است. در این مطالعه، ارزیابی مقایسه‌ای کارایی دو مدل ماشین بردار پشتیبان و شبکه عصبی مصنوعی در پیش‌بینی پارامترهای کیفی رودخانه دیناچال مورد بحث است که در آن نه پارامتر کیفی مورد پیش‌بینی قرار گرفته و توسط شاخص‌های مختلف ارزیابی شده است.

1 Artificial Neural Network

2 Support Vector Machine

مواد و روش‌ها

همانطور که در بخش پیشینه پژوهش مطرح شد، در پژوهش حاضر نه پارامتر مهم کیفی از رودخانه دیناچال مورد ارزیابی قرار گرفته است. در این خصوص از دو روش مبتنی بر یادگیری ماشین استفاده شده است. نتایج پیش‌بینی به‌وسیله داده‌های ارزیابی با یکدیگر مقایسه شد. در این بخش ضمن تشریح منطقه مطالعاتی، جمع‌آوری داده و هر دو روش شبکه عصبی مصنوعی (ANN) با کمک کدنویسی متلب و روش ماشین بردار پشتیبان (SVM) با کمک کدنویسی پایتون تشریح شده است. شکل ۱ مراحل تحقیق حاضر را نمایش می‌دهد.

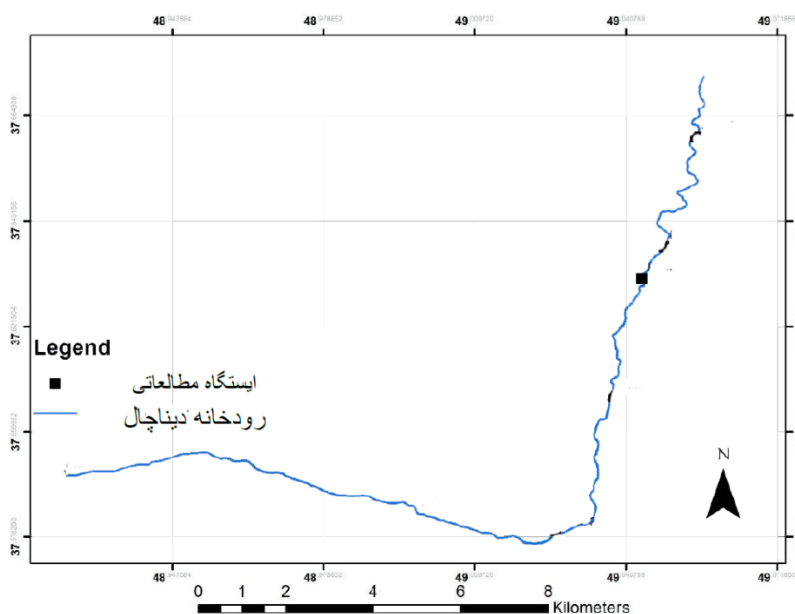


شکل ۱. مراحل تحقیق

منطقه مطالعاتی

رودخانه دیناچال یکی از رودخانه‌های دائمی غرب استان گیلان و در محدوده شهر پره‌سر از شهرستان رضوانشهر است. موقعیت جغرافیایی این شهرستان با طول شرقی ۴۸ درجه و ۴۰ دقیقه تا ۴۹ درجه و ۱۳ دقیقه و عرض شمالی بین ۳۷ درجه و ۲۵ دقیقه تا ۳۷ درجه و ۴۰ دقیقه است (امینی و همکاران، ۱۳۹۳). این رودخانه در حوضه اصلی دریای کاسپین و در حوضه درجه دو تالش و مرداب انزلی واقع شده است. محدوده مطالعاتی تحقیق حاضر، شامل بخشی از رودخانه دیناچال واقع در حدفاصل پل اتوبان رشت به آستارا و بین روستاهای دیناچال و سیمبرخاله است. حوضه آبریز رودخانه دیناچال بین حوضه‌های آبریز سفارود و ناورود واقع شده است. در این حوضه ایستگاه‌های تبخیرسنجی، باران‌سنج معمولی، باران‌سنج ذخیره‌ای ثابت و ایستگاه‌های برف‌سنجی فعال است (محمدی قلعه‌نی و همکاران، ۱۳۹۴). طبق شکل ۲ و جدول شماره ۱، موقعیت مکانی دقیق ایستگاه نمونه‌برداری نمایش داده شده است.

در مطالعه حاضر، داده‌های ماهانه کیفیت آب رودخانه دیناچال در دوره‌ی آماری موجود از سال ۱۳۸۵ تا ۱۳۹۷ از شرکت مدیریت منابع آب ایران تهیه شد و به‌منظور پیش‌بینی نه پارامتر کیفی TDS، EC، pH، یون سدیم، کلسیم، منیزیم، سولفات، کلر و کربنات با دو روش مبتنی بر یادگیری ماشین: ۱. شبکه عصبی مصنوعی (ANN) و ۲. ماشین بردار پشتیبان (SVM) مورد استفاده قرار گرفت. در این مطالعه از ۷۰ درصد داده‌ها برای آموزش مدل و از ۳۰ درصد باقیمانده برای صحت‌سنجی مدل استفاده شد. در ضمن، برای پیش‌بینی هر یک از پارامترهای معرفی شده، تعداد هشت پارامتر به‌عنوان ورودی مدل و پارامتر دیگر به‌عنوان خروجی یا هدف در نظر گرفته شد. ایستگاه مورد استفاده جهت استخراج داده‌ها و بررسی آماری آن، ایستگاه دیناچال به طول و عرض جغرافیایی $49^{\circ}27'30''$ و $37^{\circ}37'30''$ و ارتفاع ۱۰ متر از سطح آب‌های آزاد است. خصوصیات آماری داده‌های کیفی رودخانه در دوره آماری موردنظر در جدول ۲ قابل مشاهده است.



شکل ۲. نقشه موقعیت مکانی ایستگاه نمونه برداری در رودخانه دیناچال

جدول ۱. اطلاعات موقعیت مکانی ایستگاه نمونه برداری در رودخانه دیناچال

طول جغرافیایی	عرض جغرافیایی	ارتفاع از سطح دریا (متر)
۴۹° ۲' ۲۷"	۳۷° ۳۷' ۳۰"	۱۰

جدول ۲. ویژگی های آماری پارامترهای کیفی داده های ماهانه ایستگاه دیناچال در رودخانه دیناچال (۱۳۹۷-۱۳۸۵)

TDS	EC	pH	HCO ₃	Cl	SO ₄	Ca	Mg	Na	پارامتر
mg/l	ds/m	-	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	واحد
				۱۴۴					تعداد
۲۰۸/۵۹	۳۳۱/۱۵۳	۷/۷۳۸	۲/۷۸۰	۰/۲۲۷	۰/۲۹۸	۲/۳۸۱	۰/۶۵۰	۰/۲۵۲	میانگین
۴۳/۵۹	۶۹/۱۶۸	۱/۲۳۲	۰/۵۹۷	۰/۰۸۵	۰/۱۴۵	۰/۵۶۲	۰/۲۷۵	۰/۰۶۵	انحراف معیار
۹۹/۵۴	۱۵۸	۶/۷۰	۱/۰۰۰	۰/۱۰۰	۰/۰۸۰	۱/۲۰۰	۰/۱۹۰	۰/۰۴۰	کمینه
۱۸۲/۹۲	۲۹۳	۷/۱۵	۲/۵۰۰	۰/۱۹۸	۰/۲۰۸	۱/۹۹۳	۰/۴۵۰	۰/۲۲۰	%۲۵
۲۱۱/۵۲	۳۳۵/۵	۷/۵۶	۲/۸۶۵	۰/۲۰۰	۰/۲۷۰	۲/۴۳۰	۰/۶۱۰	۰/۲۶۰	%۵۰
۲۲۹/۰۰	۳۶۳/۲۵	۷/۹۱	۳/۰۶۰	۰/۲۵۰	۰/۳۳۳	۲/۷۰۰	۰/۸۱۰	۰/۲۹۰	%۷۵
۳۸۰/۰۰	۶۰۲	۱۵/۲۹	۴/۹۰۰	۰/۶۰۰	۰/۹۸۰	۴/۵۵۰	۱/۵۶۰	۰/۴۸۰	بیشینه

شبکه عصبی مصنوعی (ANN)

در این مطالعه، برای تخمین پارامترهای کیفی رودخانه دیناچال از یک شبکه عصبی مصنوعی استفاده شد. در واقع ANN شبکه‌ای از عناصر به هم پیوسته است که از مطالعات شبیه‌سازی مغز بیولوژیکی الهام گرفته شده است و می‌تواند با شناسایی الگوهای مشابه، سری زمانی را پیش بینی کند (Gorashi et al., 2012). شبکه‌های عصبی مدل‌های انعطاف‌پذیری هستند که کاربرد خود را در طیف گسترده‌ای از مسائل منابع آب نشان داده‌اند. مطابق بسیاری از مطالعات، شبکه‌های عصبی از مدل‌های کلاسیک عملکرد بهتری ارائه داده‌اند (Banejad et al., 2011). در پژوهش حاضر، از نرم‌افزار MATLAB 2019-a برای توسعه‌ی شبکه عصبی استفاده شد. یک شبکه عصبی پیش‌خور ۱ دولایه‌ای با دو تابع انتقال مختلف توسط الگوریتم لونیگ-مارکواریت ۲ آموزش داده شد؛ به طوری که از تابع سیگموئید در لایه‌ی پنهان و از تابع خطی در لایه‌ی پنهان خروجی استفاده شد. شبکه عصبی پیش‌خور از ساده‌ترین انواع شبکه عصبی است و به شبکه‌ای از لایه‌ها و نورون‌ها اطلاق

می‌شود که اطلاعات در آن فقط روبه‌جلو حرکت می‌کند و حلقه‌ای تشکیل نمی‌شود. ساختار شبکه عصبی استفاده‌شده در این مطالعه شامل هشت ورودی، ۱۰ نورون در لایه پنهان، یک نورون در لایه پنهان خروجی و یک خروجی است که همان متغیر پیش‌بینی شده است. ساختار شبکه مطابق شکل ۲ است. برای آموزش مدل از ۷۰٪ داده‌ها به‌صورت تصادفی و برای صحت سنجی و آزمودن مدل به ترتیب ۱۵٪ و ۱۵٪ داده‌ها بکار گرفته شد.

ماشین بردار پشتیبان (SVM)

مفهوم ماشین بردار پشتیبان (SVM) برای اولین بار توسط Vapnik و همکارانش در سال ۱۹۹۵ ارائه شد. کاربرد اصلی SVM در مسائل کلاسه‌بندی و پیش‌بینی داده‌ها است. در ادبیات فنی، وقتی از الگوریتم‌های SVM به‌منظور طبقه‌بندی داده‌ها استفاده شود اختصاراً SVC و زمانی که برای کاربردهای رگرسیونی استفاده شود آن را SVR نامیده‌اند. بردارهای پشتیبان در واقع نقاطی هستند که ضرایب لاگرانژ غیر صفر دارند. مفهوم تابع زیان $\epsilon - insensitive$ در SVR برای ارزیابی تابع رگرسیونی استفاده می‌شود. اصلی‌ترین هدف SVR یافتن تابع $f(x)$ است که در آن ϵ دارای بیشترین انحراف از داده‌های هدف y_i بوده و از داده‌های آموزشی به‌دست‌آمده باشد (Rohmah et al., 2021). زمانی که اسپیلون برابر با صفر شود بهترین رگرسیون به دست خواهد آمد. بالعکس، مقادیر زیاد اسپیلون بیانگر دقت پایین رگرسیون است. فرمولاسیون SVR به‌صورت تابع (۱) است.

$$f(x) = w^T \phi(x) + b \quad \text{رابطه (۱)}$$

$\phi(x)$ نتایج نگاشته شده‌ی تابع T در فضای داده‌های ورودی است. همچنین w بعد بردار وزنی و b میزان انحراف است. ضرایب w و b از طریق کمینه‌سازی تابع ریسک تعریف‌شده در معادله‌ی (۲) تخمین زده می‌شوند.

$$R = \min \|w\|^2 + C \frac{1}{l} \left(\sum_{i=1}^l L_{\epsilon}(y_i, f(x_i)) \right) \quad \text{رابطه (۲)}$$

که R تابع ریسک و C مقدار خطا است.

برای حل مسائل رگرسیونی، هنگامی که داده‌ها الگوی غیرخطی دارند از توابع کرنل استفاده می‌شود. برخی از توابع کرنل رایج عبارت‌اند از کرنل خطی^۲، چندجمله‌ای^۳ و تابع پایه شعاعی گاوسی^۴ که در این مدل از تابع کرنل خطی مطابق رابطه‌ی ۳ استفاده شد.

$$K(x_i, x) = (x_i^T \cdot x + 1)^d, \quad d = 1, 2, \dots \quad \text{رابطه (۳)}$$

که در آن d نشان‌دهنده درجه‌ی چندجمله‌ای است.

به‌منظور تشکیل مدل SVR جهت پیش‌بینی پارامترهای کیفی موردنظر، از کتابخانه scikit-learn استفاده شد (Pedregosa et al., 2021).

معیارهای ارزیابی کارایی مدل

پس از پیش‌بینی پارامترهای کیفی، کارایی مدل‌ها توسط معیارهای مختلفی از جمله^۵ MSE،^۶ RMSE و^۷ MAPE ارزیابی شد. معیار میانگین مربعات خطا MSE یک مقدار همیشه مثبت است که از مربع فاصله‌ی اقلیدسی بین مقادیر پیش‌بینی شده و داده‌های مشاهداتی به دست می‌آید. هرچه مقدار MSE کوچک‌تر باشد، دقت مدل پیش‌بینی کننده بیشتر است و داده‌های پیش‌بینی شده همخوانی بیشتری با مقادیر مشاهداتی دارند. مقدار MSE از رابطه (۴) به دست می‌آید:

$$MSE = \frac{1}{n} \sum (X - X')^2 \quad \text{رابطه (۴)}$$

که در آن n تعداد داده‌ها، X مقدار مشاهداتی و X' مقدار پیش‌بینی شده پارامتر موردنظر است. ریشه دوم میانگین مربعات خطا (RMSE) نیز مجذور MSE است که تمرکز داده‌ها پیرامون بهترین خط برازش داده‌شده را اندازه‌گیری می‌کند و از رابطه‌ی (۵) محاسبه

1 Lagrange Multiplier

2 Linear

3 Polynomial

4 Gaussian radial basis function

5 Mean Squared Error

6 Root Mean Square Error

7 Mean Absolute Percentage Error

می‌شود.

$$RMSE = \sqrt{\frac{1}{n} \sum (X - X')^2} \quad \text{رابطه (۵)}$$

علاوه بر موارد فوق، شاخص MAPE به منظور قیاس عملکرد مدل‌ها در مقیاس پارامترهای کیفی به کار گرفته شده و مطابق رابطه (۶) است.

$$MAPE = \frac{1}{n} \sum \frac{X - X'}{X} \quad \text{رابطه (۶)}$$

نتایج و بحث

مدل‌های پیش‌بینی SVR و ANN به ترتیب توسط کتابخانه پایتون SK-Learn و نرم‌افزار متلب ایجاد شدند و نه پارامتر کیفی TDS، EC، pH، HCO₃، Cl، SO₄، Ca، Mg و Na از داده‌های ماهانه رودخانه دیناچال - ایستگاه دیناچال مورد پیش‌بینی قرار گرفت. سپس عملکرد هر مدل برای هر یک از پارامترها توسط معیارهای ارزیابی RMSE، MSE و MAPE مقایسه شد. برای پیش‌بینی هر پارامتر، هشت پارامتر دیگر به عنوان ورودی مدل‌ها در نظر گرفته شد و پارامتر مورد نظر به عنوان تابع هدف به مدل معرفی شد. برای مثال به منظور پیش‌بینی پارامتر TDS، پارامترهای EC، pH، HCO₃، Cl، SO₄، Cl، SO₄، یون کلسیم، منیزیم و سدیم به عنوان ورودی و جهت آموزش مدل استفاده شد. همان‌طور که ذکر شد، در هر دو مدل ANN و SVR از ۷۰٪ داده‌ها برای آموزش مدل و ۳۰٪ داده‌ها به صورت کاملاً تصادفی برای صحت‌سنجی و آزمودن مدل در نظر گرفته شد. در جدول ۳ نتایج کارایی هر مدل برای پیش‌بینی پارامترهای فوق‌الذکر قابل مشاهده است.

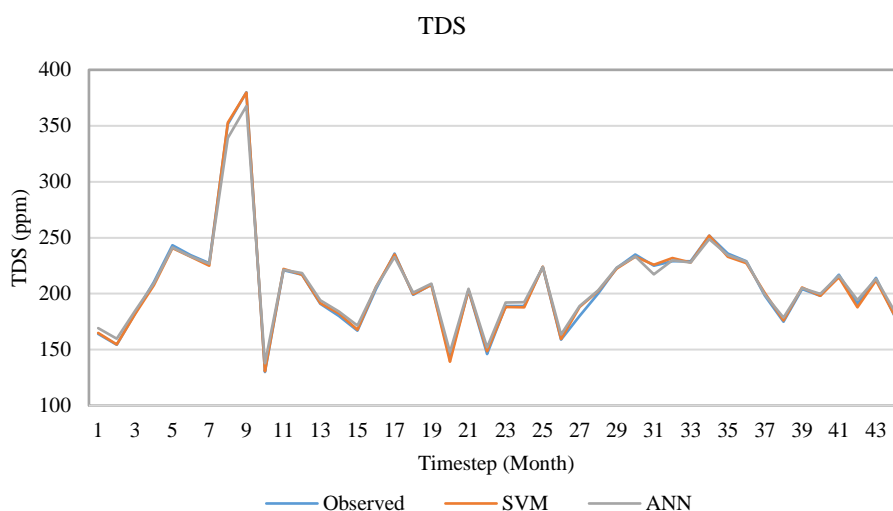
جدول ۳. قیاس عملکرد روش SVM و ANN به منظور پیش‌بینی پارامترهای کیفی رودخانه دیناچال با استفاده از معیارهای RMSE، MSE و MAPE

ماشین بردار پشتیبان (SVM)			شبکه عصبی مصنوعی (ANN)			پارامتر کیفی
MAPE	MSE	RMSE	MAPE	MSE	RMSE	
۰/۰۷۵	۴/۱۴	۲/۰۳	۰/۰۱۶	۱۹/۱	۴/۳۷	TDS
۰/۰۷۷	۸/۴۷	۲/۹۱	۰/۰۱۸	۰/۵۳۸	۰/۷۳۳	EC
۰/۱۰۴	۲/۳	۱/۵۱	۰/۱۱۷	۰/۹۴۳	۰/۹۷۱	pH
۰/۰۲۵	۰/۰۰۸	۰/۰۸۹	۰/۰۲	۰/۰۰۵۳	۰/۰۷۳	HCO ₃
۰/۲۰۱	۰/۰۰۳	۰/۰۵۵	۰/۲	۰/۰۰۳	۰/۰۵۲	Cl
۰/۰۱۸	۰/۰۰۸	۰/۰۸۷	۰/۰۱۴	۰/۰۰۲	۰/۰۴۶	SO ₄
۰/۰۲۱	۰/۰۰۵	۰/۰۶۸	۰/۰۱۷	۰/۰۰۳	۰/۰۵	Ca
۰/۰۸۲	۰/۰۰۳۹	۰/۰۶۲	۰/۰۸۲	۰/۰۰۴	۰/۰۶۳	Mg
۰/۱۴۴	۰/۰۰۲	۰/۰۴۵	۰/۰۸۹	۰/۰۰۰۷	۰/۰۲	Na

مطابق اطلاعات جدول ۳، پارامتر کل مواد جامد محلول با $RMSE=2/03$ توسط مدل SVR با دقت بهتری نسبت به شبکه عصبی مصنوعی با $RMSE=4/37$ پیش‌بینی شد. همچنین مدل SVR در پیش‌بینی یون منیزیم اندکی بهتر عمل کرد. در باقی موارد شبکه عصبی مصنوعی پیش‌بینی‌های بهتری در خصوص پارامترهای دیگر ارائه داد. اگرچه هر دو مدل از دقت مناسبی برخوردار بودند و تفاوت کارایی آن‌چنانی نداشتند. علاوه بر قیاس دو مدل توسط شاخص‌های RMSE و MSE که توضیح داده شد؛ معیار MAPE برای قیاس کارایی هر یک از مدل‌ها در پیش‌بینی پارامترهای کیفی مناسب است. مطابق نتایج به دست آمده، شبکه عصبی مصنوعی، به ترتیب در پیش‌بینی پارامترهای هدایت الکتریکی، کل مواد جامد محلول، یون کلسیم، کربنات، یون منیزیم، یون سدیم، اسیدیت، سولفات و کلر بهترین کارایی را داشته است. در مقابل، پارامتر کل مواد جامد محلول دارای بیشترین دقت توسط مدل SVR هست و پارامترهای EC، سولفات، کلسیم، کربنات، منیزیم، pH، سدیم و کلر به ترتیب دارای بهترین رتبه مطابق نتایج مدل SVR می‌باشند. بیشترین خطای تولید شده بین مقادیر مشاهداتی و پیش‌بینی شده در هر دو مدل مربوط به کلر است. همچنین EC و TDS پارامترهایی بودند نسبت به سایر پارامترهای کیفی با خطای کمتری پیش‌بینی شدند.

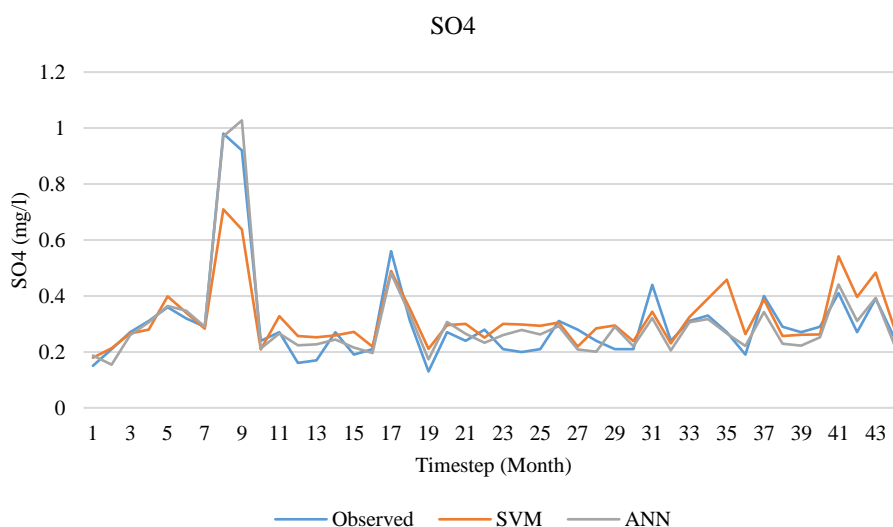
در شکل‌های ۳، ۴ و ۵، گراف‌های استنتاج شده از پیش‌بینی مدل‌ها برای پارامترهای TDS، SO₄ و Mg نشان داده شده است.

مطابق شکل ۳، هر دو مدل ANN و SVM به خوبی آموزش دیده‌اند و تطابق خوبی با داده‌های مشاهداتی از خود نشان داده‌اند، اگرچه مدل SVM در پیش بینی TDS طبق شاخص RMSE اندکی از مدل دیگر دقیق تر بوده است. در این نمودار، خطوط طوسی رنگ نشان دهنده نتایج با مدل ANN و خطوط نارنجی رنگ نشانگر داده‌های پیش بینی شده توسط مدل SVM است. همچنین خطوط آبی رنگ بیانگر داده‌های مشاهداتی است. در این نمودار محور عمودی نماینده میزان غلظت مواد جامد محلول برحسب ppm و محور افقی بیانگر زمان بر حسب ماه است.



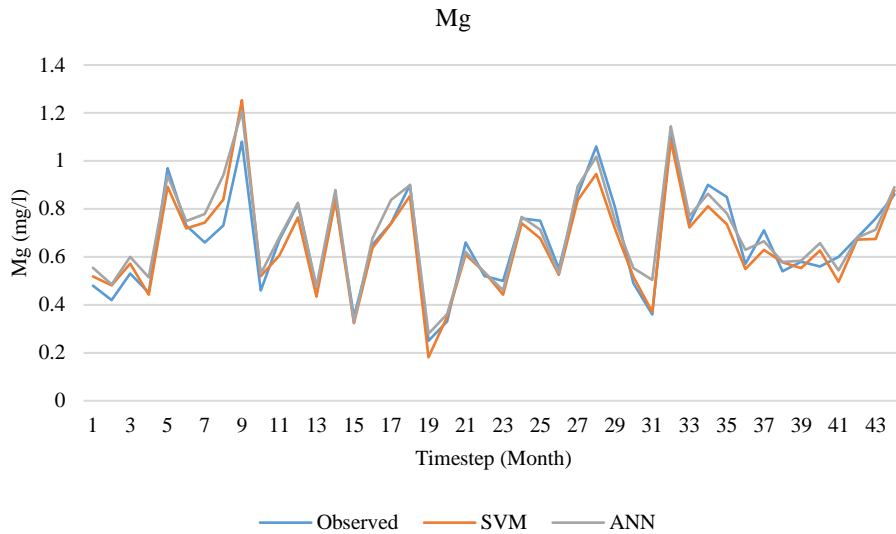
شکل ۳. نمودار مقایسه پیش‌بینی TDS توسط مدل ANN و SVM در قیاس با مقادیر مشاهداتی در دوره صحت‌سنجی

همچنین بر اساس شکل ۴، کارایی مدل‌ها را در پیش‌بینی سولفات در دوره صحت‌سنجی نشان می‌دهد که نسبت به TDS کارایی ضعیف‌تری از خود نشان داده‌اند. با این وجود، پیش‌بینی‌های استنتاج شده از مدل ANN تطابق بیشتری با داده‌های مشاهداتی دارد. همچنین در این نمودار رنگ‌های طوسی، نارنجی و آبی به ترتیب نشانگر داده‌های پیش‌بینی شده توسط مدل‌های ANN، SVM و داده‌های مشاهداتی است. محور عمودی نشان‌دهنده میزان غلظت سولفات و محور افقی زمان برحسب ماه را نشان می‌دهد.



شکل ۴. نمودار مقایسه پیش‌بینی یون سولفات توسط مدل ANN و SVM در قیاس با مقادیر مشاهداتی در دوره صحت‌سنجی

در شکل ۵ نیز، کاتیون منیزیم مورد پیش‌بینی قرار گرفته که هر دو مدل کارایی بسیار مشابه با RMSE به ترتیب برابر با ۰/۰۶۳ و



شکل ۵. نمودار مقایسه پیش‌بینی یون منیزیم توسط مدل SVM و ANN در قیاس با مقادیر مشاهداتی در دوره صحت‌سنجی

نتیجه‌گیری

در پژوهش حاضر، دو مدل مبتنی بر یادگیری ماشین، شبکه عصبی مصنوعی (ANN) و ماشین بردار پشتیبان رگرسیونی (SVR) به منظور پیش‌بینی پارامترهای کیفی رودخانه دیناچال، در دوره‌ی ۱۲ ساله منتهی به سال ۱۳۹۷، با گام‌های زمانی یک ماه طی سال‌های ۱۳۸۵ تا ۱۳۹۷ به مدل آموزش داده شد. در این مطالعه، نه پارامتر شامل pH، EC، TDS، HCO_3 ، Ca ، SO_4 ، Cl ، Mg و Na مورد بررسی و پیش‌بینی قرار گرفتند. به منظور ارزیابی کارایی مدل‌های پیش‌بینی از شاخص‌های RMSE، MSE و MAPE استفاده شد. در نهایت مطابق نتایج به دست آمده مشخص شد که در رودخانه دیناچال عملکرد مدل ماشین بردار پشتیبان در خصوص پیش‌بینی پارامترهای TDS و Mg نسبت به مدل شبکه عصبی مصنوعی مطلوب‌تر بود؛ اما مدل مبتنی بر شبکه عصبی مصنوعی در پیش‌بینی هفت پارامتر دیگر شامل pH، EC، HCO_3 ، SO_4 ، Cl ، Ca و Na عملکرد نسبتاً بهتری نسبت به مدل ماشین بردار پشتیبان از خود نشان داد. با این وجود، هر دو مدل کارایی مناسبی در پیش‌بینی پارامتر Cl نداشتند و ضریب MAPE به صورت مشابه برای هر دو مدل برابر با ۰/۲ محاسبه شد. بطور کلی عملکرد هر دو مدل در پیش‌بینی پارامترها، مناسب ارزیابی می‌شود. در نهایت پیشنهاد می‌شود در مطالعات آینده علت دقت پایین مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان در پیش‌بینی پارامتر کلر بررسی شده و راهکار مناسب ارائه شود. همچنین، کارایی مدل‌های پیشرفته‌تر مبتنی بر یادگیری عمیق در پیش‌بینی پارامترهای کیفی می‌تواند مورد بررسی و مطالعه قرار بگیرد.

تشکر و قدردانی

بدین وسیله از دانشگاه تهران و شرکت آب منطقه‌ای استان گیلان به دلیل تأمین امکانات لازم جهت انجام این تحقیق و تهیه مقالات مربوطه تشکر و قدردانی می‌شود.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

مراجع

- احمدی، فرشاد؛ رادمنش، فریدون و میرعباسی نجف‌آبادی، رسول (۱۳۹۳). مقایسه روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوزچای). *نشریه آب و خاک (علوم و صنایع کشاورزی)*، ۱۱۷۱-۱۱۶۲ و ۲۸(۶)
- امینی، عباس؛ نوری، سید هدایت اله و اصلانی سنگده، بیتا. (۱۳۹۳). ارزیابی و سنجش پایداری زراعت برنج با استفاده از روش‌های تصمیم‌گیری چند معیاره (مورد مطالعه: شهرستان رضوانشهر). *علوم ترویج و آموزش کشاورزی ایران*، ۱۲۶-۱۰۱(۱)۱۱.
- ایرجی، مریم؛ موحدی نائینی، سید علیرضا؛ کمکی، چوقی بایرام؛ ابراهیمی، سهیلا و یغمایی، بامشاد. (۱۴۰۲). ارزیابی پارامترهای موثر جهت پیش‌بینی

- عیار پتاسیم شورابه با استفاده از الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی (مطالعه موردی: پلایای شهرستان خورو بیابانک، استان اصفهان). *مجله تحقیقات آب و خاک ایران*.
- جمالی، بهزاد و ابراهیمی، کیومرث. (۱۳۹۰). پیش‌بینی سری‌های زمانی کیفیت آب رودخانه سفیدرود با استفاده از مدل‌های خطی تصادفی. *مجله تحقیقات مهندسی کشاورزی*، ۴۴-۳۱ و ۱۲(۳).
- حسین‌پناهی، بهاره؛ نیک‌مهر، سامان و ابراهیمی، کیومرث. (۱۴۰۰). مقایسه مدل‌های ماشین بردار و شبکه عصبی تابع پایه شعاعی در پیش‌بینی کیفیت آب سیمینه‌رود. *مدیریت آب و آبیاری*، ۴۱۹-۴۰۹ و ۱۱(۳).
- روشنگر، کیومرث و داودی، سینا. (۱۴۰۱). مقایسه عملکرد روش‌های یادگیری عمیق و یادگیری ماشین در پیش‌بینی میزان اکسیژن محلول. *مجله تحقیقات آب و خاک ایران*، ۱۹۰۰-۱۸۸۵(۸)۵۳.
- ستاری، محمدتقی؛ عباسقلی نایب زاد، مهدی و میرعباسی نجف‌آبادی، رسول. (۱۳۹۳). پیش‌بینی کیفیت آب‌های سطحی با استفاده از روش درخت تصمیم. *نشریه علمی پژوهشی مهندسی آبیاری و آب ایران*، ۸۸-۸۶(۳)۴.
- شاهی‌نژاد، بابک؛ ایزدی، زهره و جوادی، بهزاد. (۱۴۰۰). ارزیابی مدل Qual2Kw در شبیه‌سازی کیفی رودخانه خرم‌آباد. *نشریه هیدروژئومورفولوژی*، دوره ۸، شماره ۲۶.
- عباسی، عباس؛ خلیلی، کیوان؛ بهمنش، جواد و شیرزاد، اکبر. (۱۳۹۹). کاربرد روش ماشین بردار پشتیبان و شبکه بیزین در پیش‌بینی خشک‌سالی کشاورزی. *مهندسی و مدیریت آبخیز* ۱۲۴-۱۰۷ و ۱۲(۱).
- محمدی قلعه‌نی، مهدی؛ ابراهیمی، کیومرث و امید، محمدحسین. (۱۳۹۴). برآورد ضریب هوادهی در مدیریت کیفی آب رودخانه دیناچال، با استفاده از روابط تجربی و روش‌های عددی. *مدیریت آب و آبیاری*، ۷۹-۶۹(۱)۵.
- ویسی، محمد؛ محمدی روزبهانی، مریم و ظهراپی، نرگس. (۱۳۹۶). بررسی کیفیت آب رودخانه شاوور با استفاده از شاخص‌های کیفی آب (مطالعه موردی: روستای حمزه تا سد شاوور). *فصلنامه علوم و تکنولوژی محیط‌زیست*، ۹۳-۸۱(۵)۲۳.

REFERENCES

- Abbasi, A., Khalili, K., Bahmanesh, J & Shirzad, A. (2020). Application of Support Vector Machine and Bayesian Network for Agricultural Drought Prediction. *Journal of Watershed Engineering and Management*, 107-124 and (1)12. (In Persian)
- Abobakr Yahya, A. S. Ahmed, A. N. Binti Othman, F. Ibrahim, R. K. Afan, H. A. El-Shafie, A. & Elshafie, A. (2019). Water Quality Prediction Model-Based Support Vector Machine Model for ungauged River Catchment under Dual Scenarios. *Water*, 11(6), 1231.
- Ahmadi, F., Radmanesh, F & Mirabbasi Najaf Abadi, R. (2015). Comparison between Genetic Programming and Support Vector Machine Methods for Daily River Flow Forecasting (Case Study: Barandoozchay River). *Journal of Water and Soil*, 1162-1171 and 28 (6). (In Persian)
- Amini, A., Nouri, HA & Aslani Sangdeh, B. (2015). Evaluation of Rice Production Sustainability Using Multi-Criteria Decision-Making Methods: The Case of Razvanshahr County. *Agricultural Education and Extension Sciences of Iran*, 126-101(1)11. (In Persian)
- Banejad, H. and E. Olyaie (2011). "Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction—A Case Study." *Journal of American Science Journal of American Science*, 77: 60-65.
- Chia, S. L. Chia, M. Y. Koo, C. H. & Huang, Y. F. (2022). Integration of Advanced Optimization Algorithms into Least-Square Support Vector Machine (LSSVM) for Water Quality Index Prediction. *Water Supply*, 22(2), 1951-1963.
- Cortes, C. and V. Vapnik (1995). "Support-vector networks." *Machine Learning* 20(3): 273-297.
- Gorashi, F. & Abdullah, A. (2012). Prediction of Water Quality Index Using Back Propagation Network Algorithm Case Study: Gombak River. *Journal of Engineering Science and Technology*, 7(4), 447-461.
- Hosseinpanahi, B., Nikmehr, S & Ebrahimi, K. (2021). Comparison of the Support Vector Machine Models and Radial Function Neural Network in Predicting SiminehRood River Water Quality Iran. *Water and Irrigation Management*, 409-419 and (3)11. (In Persian)
- Huang, M. Tian, D. Liu, H. Zhang, C. Yi, X. Cai, J., & Ying, G. (2018). A Hybrid Fuzzy Wavelet Neural Network Model with Self-Adapted Fuzzy c-means Clustering and Genetic Algorithm for Water Quality Prediction in Rivers. *Complexity*, 2018, 1-11.
- Iraji, M., Movahedi Naeini, SA., Komaki, CB., Ebrahimi, S & Yaghmaei, B. (2023). Evaluation of the Effective Parameters for Predicting the Potassium Grade of Saline Water by Using Support Vector Machine and Random Forest Algorithms (Case Study: Playa of Khor and Biabank Area City, Isfahan



- Province). *Iran water and soil research*. (In Persian)
- Jamali, B & Ebrahimi, K. (2011). Time Series Forecasting of Sefidrood River Water Quality Using Linear Stochastic Models. *Journal of Agricultural Engineering Research*, 31-44 and (3)12. (In Persian)
- Li, Q., Yang, Y., Yang, L., & Wang, Y. (2023). Comparative Analysis of Water Quality Prediction Performance Based on LSTM in the Haihe River Basin, China. *Environmental Science and Pollution Research*, 30(3), 7498-7509.
- Mahapatra, S. S., Nanda, S. K., & Panigrahy, B. K. (2011). A Cascaded Fuzzy Inference System for Indian River Water Quality Prediction. *Advances in Engineering Software*, 42(10), 787-796.
- Mohammadi Ghaleni, M., Ebrahimi, K & Omid, MH. (2015). Estimation of Reaeration Rate Coefficient on Dinachal River Numerical Methods in Water Quality Management. *Water and Irrigation Management*, 79-69(1)5. (In Persian)
- Najwa Mohd Rizal, N., Hayder, G., Mnzool, M., Elnaim, B. M., Mohammed, A. O. Y., & Khayyat, M. M. (2022). Comparison between Regression Models, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in River Water Quality Prediction. *Processes*, 10(8), 1652.
- Pedregosa, F., et al. (2012). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12.
- Rohmah, M. F., et al. (2021). "Comparison Four Kernels of SVR to Predict Consumer Price Index." *Journal of Physics: Conference Series* 1737(1): 012018.
- Roshangar, K & Davoudi, S. (2022). Comparing the Performance of Deep Learning and Machine Learning Methods in Predicting Dissolved Oxygen Content. *Iran soil and water research*, 1885-1900(8)53. (In Persian)
- Sakaa, B., Elbeltagi, A., Boudibi, S., Chaffai, H., Islam, A. R. M. T., Kulimushi, L. C., & Wong, Y. J. (2022). Water Quality Index Modeling Using Random Forest and Improved SMO Algorithm for Support Vector Machine in Saf-Saf River Basin. *Environmental Science and Pollution Research*, 29(32), 48491-48508.
- Sattari, M., Abbasgoli Naebzad, M & Mirabbasi Najafabadi. (2014). Surface Water Quality Prediction Using Decision Tree Method. *Iranian Journal of Irrigation and Water Engineering*, 88-86(3)4. (In Persian)
- Shahinejad, B., Izadi, Z & Javadi, B (2021). Evaluation of Qual2Kw Model in Qualitative Simulation of Khorramabad River. *Hydrogeomorphology*. Volume 8, Number 26. (In Persian)
- Sillberg, C.V., Kullavanijaya, P., & Chavalparit, O. (2021). Water Quality Classification by Integration of Attribute-Realization and Support Vector Machine for the Chao Phraya River. *Journal of Ecological Engineering*, 22 (9), 70-86
- Veisi, M & Mohammadi Rouzbahani, M. (2021). Survey on Shavoor River Using Water Quality Indexes (Hamzeh village to Shavoor dam). *Environmental Science and Technology Quarterly*, 23-81(5)-93. (In Persian)
- Yan, J., Liu, J., Yu, Y., & Xu, H. (2021). Water Quality Prediction in the Luan River based on 1-drcnn and bigru Hybrid Neural Network Model. *Water*, 13 (9), 1273.

Comparison of artificial neural network methods and support vector machine in predicting water quality parameters of Dinachal river, IRAN

EXTENDED ABSTRACT

Introduction

Rivers play a crucial role as vital sources of drinking water and agricultural irrigation, supporting both human health and the environment. However, the presence of various pollutants from agricultural, industrial, and urban sources can turn these invaluable resources into vulnerable points of contamination. Therefore, accurately predicting water quality is essential for effectively monitoring river ecosystems and ensuring their reliability. Unfortunately, traditional forecasting models often struggle to capture the complex and dynamic nature of water quality changes, characterized by non-linear and non-stationary patterns. Nonetheless, recent advancements in artificial neural networks have revolutionized the field of water quality prediction, sparking new discussions and possibilities in this area.

Materials and Methods

This study focuses on the evaluation and prediction of qualitative parameters in the Dinachal River, a vital river in the Gilan province. Two models, namely artificial neural network and regression support vector machine, were utilized for predicting nine qualitative parameters including total dissolved solids, electrical conductivity, acidity, Cl, SO₄, HCO₃, Ca, Mg, and Na. The predictions were made on a monthly basis from the years 2006 to 2, and the performance of the models was assessed using RMSE, MSE, and MAPE statistics.

The artificial neural network was employed to estimate the quality parameters of the Dinachal River. Inspired by biological brain simulation studies, the ANN is a network of interconnected elements capable of identifying similar patterns and predicting time series data. Neural networks have demonstrated their flexibility and applicability in various water resources problems, often outperforming traditional models. For this study, MATLAB 2019a software was employed to develop the neural network. The model was trained using 70% of the data randomly selected, while 15% and 15% were used for validation and testing purposes, respectively.

In the technical literature, support vector machine algorithms are abbreviated as SVC for data classification and SVR for regression applications. Support vectors refer to points with non-zero Lagrange coefficients, and the loss function in SVR is used to evaluate the regression function. The primary objective of SVR is to find the function $f(x)$ that deviates the most from the target data, as obtained from the training data. When the deviation is zero, the best regression is achieved, while larger values indicate lower regression accuracy.

Results

The prediction models SVR and ANN were developed using the SK-Learn Python library and MATLAB software, respectively. These models were utilized to predict nine qualitative parameters (TDS, EC, pH, HCO₃, Cl, SO₄, Ca, Mg, and Na) based on monthly data from the Dinachal River - Dinachal Station. The performance of each model for each parameter was then compared using evaluation indices such as RMSE, MSE, and MAPE.

To predict each parameter, eight other parameters were used as inputs for the models, with the desired parameter being treated as the objective function. For instance, when predicting the TDS parameter, inputs such as EC, pH, HCO₃, Cl, SO₄, calcium, magnesium, and sodium ions were considered during model training. In both the ANN and SVR models, 70% of the data was allocated for training the models, while the remaining 30% was randomly selected for validation and testing purposes.

Conclusion:

Based on the results obtained, the SVR model demonstrated superior performance in predicting total dissolved solids (with an RMSE of 2.03) and magnesium concentration (with an RMSE of 0.062) compared to the artificial neural network model. Conversely, the artificial neural network model exhibited relatively better success in predicting the remaining parameters. Nevertheless, both models were deemed suitable for predicting the quality parameters of the Dinachal River.

Furthermore, the SVR model achieved the best performance in predicting total dissolved solids and electrical conductivity with a MAPE coefficient of 0.007, while the artificial neural network model achieved a MAPE coefficient of 0.001 for the same parameters. It is worth mentioning that both models showed weaker performance in predicting the chlorine parameter, despite having RMSE values of 0.055 and 0.052 for SVR and ANN models, respectively.

The methods employed in this study have proven effective in predicting the water quality of the Dinachal



River overall. However, it is recommended that future studies investigate the reasons behind the lower accuracy of the artificial neural network and support vector machine models in predicting the chlorine parameter, and propose appropriate solutions. Additionally, exploring the performance of more advanced models based on deep learning for predicting quality parameters could be a valuable avenue for further research.

Keywords: ANN, Dinachal River, SVR, Water Quality Prediction.