



Available online: <http://ijhe.tums.ac.ir>

مقاله پژوهشی

## مدلسازی پارامترهای کیفی (هدایت الکتریکی و کل مواد جامد محلول) رودخانه کارون در ایستگاه‌های ملاثانی، اهواز و فارسیات با بهره‌گیری از روش‌های مبتنی بر داده‌کاوی

علیرضا محقق، مهدی ولیخان انارکی، سعید فرزین\*  
گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران

اطلاعات مقاله:	چکیده
تاریخ دریافت: ۹۸/۱۲/۱۷	زمینه و هدف: رودخانه‌ها از مهمترین منابع آب هستند که آلودگی این منابع می‌تواند
تاریخ ویرایش: ۹۹/۰۳/۱۰	تاثیر مستقیم بر زندگی انسان‌ها و محیط زیست داشته باشد. در پژوهش حاضر، پارامترهای
تاریخ پذیرش: ۹۹/۰۳/۱۳	کیفی EC و TDS رودخانه کارون در سه ایستگاه آب‌سنجی ملاثانی، اهواز و فارسیات با
تاریخ انتشار: ۹۹/۰۳/۳۱	استفاده از الگوریتم‌های مبتنی بر داده‌کاوی شامل LSSVM، ANFIS و ANN مدل‌سازی
	شده است.
	<b>روش بررسی:</b> هشت سری داده مختلف شامل ترکیب $K^{+1}$ ، $Mg^{+2}$ ، $Na^{+1}$ ، $Ca^{+2}$ ، $Cl^{-1}$ ، $SO_4$ ، $HCO_3$ ، $CO_3$ با دبی جریان (Q) به‌عنوان ورودی الگوریتم‌ها، به روش کالیبراسیون
	غیرتصادفی و تصادفی انتخاب شده است. نتایج دو روش، مورد مقایسه قرار گرفته است. در
	گام بعدی، مدل‌سازی EC و TDS براساس چهار پارامتر $Ca^{+2}$ ، $Cl^{-1}$ ، $Na^{+1}$ و Q با تاخیرهای
	زمانی صفر تا سه ماهه انجام شده است.
	<b>یافته‌ها:</b> پارامترهای $Ca^{+2}$ و $Cl^{-1}$ ، $Na^{+1}$ دارای بیشترین تاثیر در مدل‌سازی پارامترهای EC
	و TDS هستند. الگوریتم LSSVM در مدل‌سازی هر دو پارامتر EC و TDS دارای بیشترین
	دقت است. در میان ایستگاه‌های بررسی شده، بیشترین دقت برای مدل‌سازی EC، مربوط
	به ایستگاه اهواز و برای TDS، مربوط به ایستگاه ملاثانی است که به ترتیب ۱۶ و ۳۶
	درصد ضریب تعیین بیشتری داشته‌اند. الگوریتم LSSVM در مدل‌سازی نوسانات و حداکثر
	پارامترهای EC و TDS در طول تغییرات زمان نیز، بیشترین دقت را دارد.
	<b>نتیجه‌گیری:</b> روش‌ها و مدل‌های استفاده شده در پژوهش حاضر به‌ویژه الگوریتم
	LSSVM می‌توانند ابزار مفید تصمیم‌گیری برای پیش‌بینی و مدیریت کیفی رودخانه‌ها
	از جمله رودخانه‌های حوضه آبریز کارون باشند. نتایج مدل‌سازی پارامترهای کیفی رودخانه‌ها
	با بهره‌گیری از هر دو روش کالیبراسیون غیرتصادفی و تصادفی، قابل اعتماد و استفاده است.
	البته دقت روش کالیبراسیون تصادفی، اندکی بیشتر است.
	<b>واژگان کلیدی:</b> شوری آب، داده‌کاوی، حوضه کارون، کالیبراسیون تصادفی، کالیبراسیون غیرتصادفی
	<b>پست الکترونیکی نویسنده مسئول:</b> saeed.farzin@semnan.ac.ir

## مقدمه

مدلسازی کیفیت آب یکی از مهمترین امور در مدیریت منابع آب و تولید آب شیرین مورد نیاز مناطق شهری، کشاورزی و صنعتی است. رودخانه‌ها از مهمترین منابع آب هستند و آلودگی این منابع می‌تواند تاثیر مستقیم بر توسعه جمعیت و اقتصاد مناطق اطراف آنها داشته باشد. رودخانه‌ها در مسیر خود محل ورود منابع مختلف آلودگی هستند. یکی از این منابع آلودگی رواناب زمین‌های کشاورزی است. فاضلاب‌های شهری، فاضلاب‌های صنعتی، توسعه مناطق شهری، زهکشی معادن نیز از دیگر موارد ایجاد آلودگی در رودخانه‌ها هستند. یکی از راهکارهای ضروری برای ارائه راه‌حل‌های مناسب جهت کنترل و کاهش این آلودگی‌ها، مدلسازی کیفیت آب است.

از عوامل تاثیرگذار بر کیفیت آب میزان شوری آن است که توسط هدایت الکتریکی (EC) و مجموع مواد محلول ((Total Dissolved Solids (TDS)) اندازه‌گیری می‌شود. در سال‌های اخیر، مطالعات بسیاری در مورد مدلسازی EC و TDS رودخانه‌های مناطق مختلف با استفاده از روش‌های مبتنی بر داده‌کاوی انجام شده است. اصولاً، اهمیت مدلسازی در ایجاد آگاهی از اثر تغییرات سری متغیرهای ورودی در متغیر هدف است. در مطالعات مختلف، با مدلسازی پارامترهای شوری آب مانند EC و TDS با استفاده از  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $Na^{+1}$  می‌توان اثر تغییرات این پارامترها را بر EC و TDS بررسی کرد. به عنوان مثال در احداث یک کارخانه، مشخص است که چه مقدار از پارامترهای آلوده کننده مانند  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $Na^{+1}$  و ... تولید و وارد رودخانه می‌شوند. در چنین شرایطی با استفاده از مدلسازی EC و TDS می‌توان متوجه شد که تغییر پارامترهای مذکور چه وضعیتی را از نظر شوری آب EC و TDS ایجاد می‌کنند. بنابراین با استفاده از مدلسازی پارامترهای مذکور می‌توان تاثیر یک آلودگی محتمل را بر شوری آب رودخانه سنجید. این موضوع از اهمیت بالایی برخوردار است زیرا افزایش شوری آب می‌تواند سبب آثار

زیانبار بر زندگی انسان، کاهش عملکرد تولید محصولات زراعی استراتژیک مانند گندم و نیشکر، و آسیب به باغات از جمله نخلستان‌های خرما می‌گردد. همچنین در صورت داشتن تغییرات شاخص‌های مانند  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $Na^{+1}$  و ... (که با توجه به تعریف سناریوها برای توسعه فعالیت‌های صنعتی، کشاورزی و شهری در آینده مشخص می‌شود) می‌توان تغییرات در پارامترهای EC و TDS را بدون نیاز به اندازه‌گیری‌های میدانی سنجید.

با توجه به اهمیت موضوع، تاکنون مطالعات چندی صورت گرفته است Orouji و همکاران (۱) مدلسازی EC و TDS در آستانه سفیدرود با به‌کارگیری دو روش سیستم استنتاج تطبیقی فازی-عصبی (Adaptive Neuro Fuzzy Inference System (ANFIS)) و برنامه‌ریزی ژنتیک (Gene Expression Programming (GEP)) انجام دادند. نتایج بیانگر عملکرد مطلوب هر دو روش بود. Ravansalar و همکار (۲) با استفاده از شبکه‌های عصبی مصنوعی (Artificial Neural Network (ANN)) و هیبرید ANN و آنالیز موجک (Wavelet\_ANN) و هیبرید (WT\_ANN) به مدلسازی EC رودخانه اسی در ترکیه پرداختند. نتایج نشانگر دقت بالاتر WT\_ANN بود. Barzegar و همکاران (۳) روش‌های ANN، ANFIS و هیبرید آنها با آنالیز موجک را جهت شبیه‌سازی EC رودخانه آجی-چای در شمال ایران به‌کار بردند که نتایج نشان از برتری روش ANFIS بود. Haghiabi و همکاران (۴) پارامترهای کیفیت آب رودخانه تیره (جنوب ایران) شامل TDS و EC را با استفاده از ANN، شبکه‌های عصبی مصنوعی گروهی (Group Method of Data Handling (GMDH)) و ماشین بردار پشتیبان (Support Vector Machine (SVM)) و ماشین مدل کردند. نتایج نشان‌دهنده برتری SVM بود. در پژوهش Montaseri و همکاران (۵) به منظور مدلسازی TDS براساس EC، Na و Cl، از ANN، مدل ANFIS با روش‌های خوشه‌بندی مختلف، برنامه‌ریزی ژنتیک GEP و هیبرید روش‌های مذکور با

شمار می‌رود که تاثیر قابل توجهی در رشد اقتصاد کشور دارد. به همین دلیل این حوضه مورد توجه محققین مختلف قرار گرفته است. Naddafi و همکاران (۱۲) با استفاده از توزیع‌های آماری مختلف، روند کیفیت آب در دو ایستگاه گتوند و خرمشهر رودخانه کارون را بررسی کردند که پارامترهای TDS و EC دارای روند مثبت بودند. Musavi-Jahromi و همکار (۱۳) به مدلسازی پارامترهای کیفیت آب با استفاده از ANN و دبی جریان به عنوان ورودی پرداختند. نتایج نشان داد، عملکرد ANN نسبتاً خوب است. در پژوهش Salami و همکاران (۱۴) پارامترهای کیفیت آب رودخانه کارون با استفاده از ANN و روش‌های ریاضی شبیه‌سازی شد که نتایج نسبتاً قابل قبول حاصل شد. Ehteram و همکاران (۱۵) با استفاده از مدل IHACRES و الگوریتم خفاش به بهینه‌سازی بهره‌برداری از مخزن در زیرحوضه دز واقع در حوضه کارون پرداختند. Valikhan-Anaraki و همکاران (۱۶) جهت ریزمقیاس‌نمایی بارش، مطالعاتی انجام داده‌اند. با عنایت به جمع‌بندی مطالعات پیشین می‌توان دریافت که اصولاً برنامه‌ریزی به منظور توسعه صنعتی، کشاورزی، و شهری نیاز به داشتن دانش و آگاهی از تاثیر پارامترهای مختلف به یکدیگر است. لذا، توسعه مدل‌هایی که تاثیر ورود عوامل آلوده کننده (در اثر توسعه صنعتی، شهری و کشاورزی و تغییرات کاربری اراضی) بر پارامترهای کیفی آب مانند EC و TDS را ارزیابی کند از اهمیت بالایی برخوردار است. با وجود اهمیت بالای رودخانه کارون در بحث مدیریت منابع آب ایران و به ویژه استان خوزستان، تاکنون مطالعات جامعی در خصوص مدلسازی کیفیت آب این رودخانه با استفاده از روش‌های مبتنی بر هوش مصنوعی صورت نگرفته است. از اینرو، در پژوهش حاضر سعی شده است که پارامترهای کیفی EC و TDS آب رودخانه کارون با استفاده از روش‌های داده‌کاوی به‌روز و توانمند، مدلسازی و تحلیل شود. به منظور نیل به اهداف فوق، از روش‌های داده‌کاوی شامل

WT، استفاده شد. در این پژوهش که مطالعه موردی رودخانه نالوچای (شمال غربی ایران)، تجن (شمال ایران)، زاینده‌رود (مرکز ایران) و حله (جنوب عراق) بود، برتری WT\_GEP ثابت شد. Salarijaz و همکار (۶) برای مدلسازی EC در ایستگاه‌های زرین گل و رامیان در شمال ایران، از روش رگرسیون ساده بهبود یافته استفاده کردند. در این روش تنها از دبی جریان به عنوان ورودی مدل استفاده شد و نتایج حاکی از عملکرد قابل قبول روش بهبود یافته بود. در پژوهش AI-Mukhtar و همکار (۷) از سه روش رگرسیون خطی، ANN و ANFIS جهت مدلسازی TDS و EC در رودخانه ابو-زاری مارش در جنوب عراق استفاده کردند. نتایج نشان داد که نیترات ( $\text{NO}_3^-$ )، منیزیم ( $\text{Mg}^{+2}$ )، کلسیم ( $\text{Ca}^{+2}$ )، سولفات ( $\text{SO}_4$ ) و کلراید ( $\text{Cl}^{-1}$ ) در مدلسازی TDS و کلسیم ( $\text{Ca}^{+2}$ )، منیزیم ( $\text{Mg}^{+2}$ )، سولفات ( $\text{SO}_4$ ) و کلراید ( $\text{Cl}^{-1}$ ) در مدلسازی EC دارای بیشترین تاثیر بودند. Singh (۸) الگوریتم‌های ANN، GP و SVM را برای مدلسازی نرخ جذب سدیم (Sodium Absorption Ratio (SAR)) در ایران به کار برد. نتایج حاکی از دقت بیشتر ANN و SVM بود. Rajaei و همکاران (۹) از الگوریتم‌های رگرسیون خطی چند متغیره، ANN و WT\_ANN جهت پیش‌بینی  $\text{NO}_3$  و DO در رودخانه کرج استفاده کردند. نتایج بیانگر برتری WT\_ANN بود. Banejad و همکاران (۱۰) از الگوریتم‌های ANN و WT\_ANN برای تخمین برخی پارامترهای کیفی آب مانند EC و TDS استفاده کردند. نتایج حاکی از خطای کمتر WT\_ANN بود. Soleimanpour و همکاران (۱۱) با استفاده از روش داده کاوی درخت تصمیم طبقه‌بندی و رگرسیون (Classification and Regression Tree (CART)) موثرترین پارامترهای کیفیت آب آشامیدنی را در دشت کازرون تعیین کردند. در این پژوهش دو عامل TDS و  $\text{Ca}^{+2}$  بیشترین تاثیر را بر کیفیت آب آشامیدنی داشتند. حوضه کارون یکی از مهمترین منابع آب سطحی ایران به

شکل ۱ موقعیت سه ایستگاه آب‌سنجی بررسی شده در حوضه کارون را نشان می‌دهد.

### مواد و روش‌ها

#### - داده‌های استفاده شده

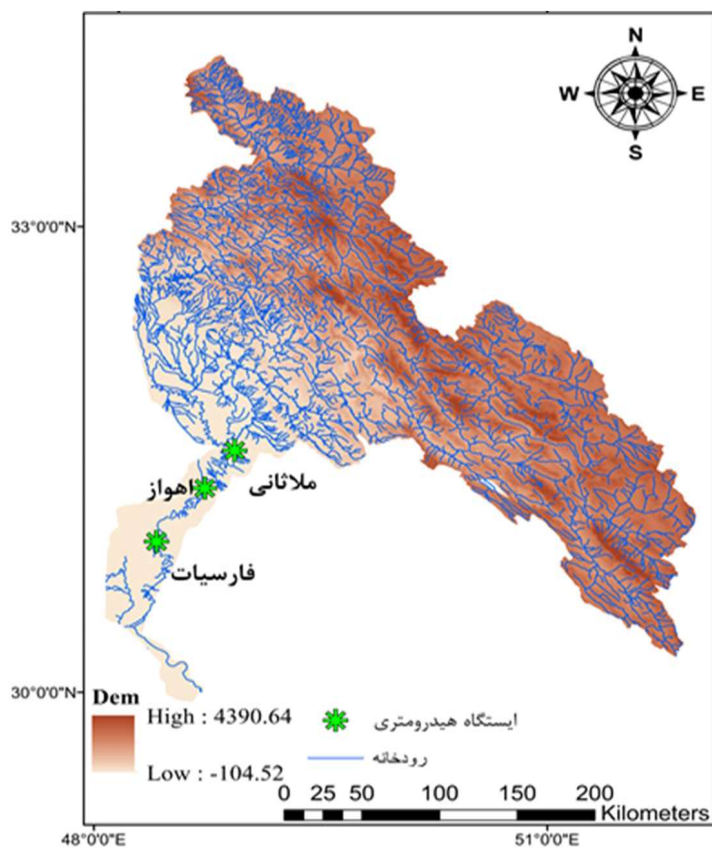
در پژوهش حاضر از پارامترهای کیفی آب شامل  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $SO_4$ ،  $HCO_3$ ،  $CO_3$ ،  $K^{+1}$ ،  $Mg^{+2}$ ،  $Na^{+1}$  و دبی جریان (Q) به منظور مدلسازی دو پارامتر مربوط به شوری آب شامل EC و TDS استفاده شده است. این داده‌ها از شرکت آب منطقه‌ای خوزستان دریافت و جهت مدلسازی با روش‌های داده‌کاوی آماده شده‌اند. در پژوهش حاضر، مانند مطالعات Ghorbani و همکاران (۱۷)، Banejad و همکاران (۱۰)، Rajaei و همکاران (۹)، Valikhan Anaraki و همکاران (۱۸) تحلیل‌های آماری براساس معیارهای میانگین، انحراف معیار، ضریب تغییرات و ضریب چولگی انجام شده است. جدول ۱ مشخصات آماری داده‌های ورودی و خروجی را براساس معیارهای میانگین، انحراف معیار، ضریب تغییرات و ضریب چولگی ارائه می‌دهد. طول دوره آماری برای داده‌های ایستگاه ملاثانی، اهواز و فارسیات به ترتیب از سال ۱۳۴۸ تا ۱۳۹۱ با ۵۰۴ داده، از سال ۱۳۴۶ تا ۱۳۹۰ با ۵۲۹ داده، و از ۱۳۶۲ تا ۱۳۹۰ با ۳۳۳ داده است. لازم به ذکر است که طول دوره‌های آماری در نظر گرفته شده، بر مبنای حداکثر تعداد داده در دسترس است.

#### - مدلسازی کیفیت آب رودخانه

در پژوهش حاضر جهت مدلسازی کیفیت آب در ایستگاه ملاثانی، اهواز و فارسیات از الگوریتم‌های LSSVM، ANFIS و ANN استفاده شده است. به این صورت که ابتدا ۷۰ درصد داده‌ها برای دوره آموزش و ۳۰ درصد داده‌ها برای دوره آزمون در نظر گرفته می‌شوند (۱۸). این روش در سال‌های اخیر در مطالعات معتبر داخلی و خارجی بسیاری (۲۱-۱۹) استفاده شده است. سپس شوری آب براساس داده‌های دوره آموزش و آزمون مدلسازی می‌شود. به این

(Least Square Support Vector Machine) LSSVM و ANFIS بهره گرفته شده است. همچنین، با استفاده از روش‌های مذکور تاثیر پارامترهای مختلف آلوده کننده بر شوری آب ارزیابی شده است. لازم به ذکر است که روش LSSVM با استفاده از تابع کرنل و تبدیل مسائل غیرخطی به خطی و روش ANFIS با ترکیب توانایی یادگیری ANN و قدرت استدلال فازی، از توانایی بالایی در حل مسائل غیرخطی و پیچیده از جمله مدلسازی کیفیت آب برخوردارند. در ابتدا، با استفاده از آنالیز حساسیت پارامترهای که بیشترین تاثیر را بر EC و TDS آب دارند مشخص شده است. در ادامه، جهت ارزیابی عملکرد و کارایی دو روش مذکور، نتایج حاصل، با نتایج ANN مقایسه گردیده است. سپس، براساس الگوریتم منتخب (از نظر بیشترین دقت) و پارامترهای مشخص شده در بخش آنالیز حساسیت مدلسازی EC و TDS انجام شده است. همچنین، مدلسازی EC و TDS براساس دو روش کالیبراسیون غیرتصادفی و تصادفی مورد مقایسه قرار گرفته است.

مطالعه موردی در پژوهش حاضر سه ایستگاه ملاثانی، اهواز و فارسیات واقع در حوضه کارون بزرگ است. دلیل انتخاب حوضه کارون، وجود یکی از مهمترین شریان‌های حیاتی کشور به نام رودخانه کارون است. این حوضه مساحتی بالغ بر  $67257 \text{ km}^2$  را زهکشی می‌کند. حوضه کارون بزرگ، بیشتر مساحت استان چهارمحال بختیاری، بخش‌های بزرگی از استان‌های لرستان و خوزستان و بخش‌های از استان‌های اصفهان، مرکزی و کهگیلویه و بویراحمد را تشکیل می‌دهد. جریان زهکش شده از مناطق مذکور از ایستگاه‌های ملاثانی، اهواز و فارسیات می‌گذرد. بنابراین، در پژوهش حاضر، به منظور مدلسازی کیفیت آب، سه ایستگاه ملاثانی (واقع در بالادست اهواز)، اهواز و فارسیات (واقع در پایین دست اهواز) در نظر گرفته شده است که معیشت و زندگی افراد بسیاری (حدود دو میلیون نفر در جنوب غربی کشور ایران) را تحت تاثیر قرار می‌دهند.



شکل ۱- موقعیت سه ایستگاه آب‌سنجی بررسی شده در حوضه کارون

جدول ۱- مشخصات آماری داده‌های ورودی و خروجی

نوع داده	ضریب چولگی	ضریب تغییرات	انحراف معیار	میانگین	معیار
ورودی	۲/۷۶	۰/۹۹	۶۲۷/۸۷	۶۱۸/۸۷	Q
ورودی	۱/۶۰	۰/۰۱	۱/۴۳	۴/۳۷	Ca <sup>+2</sup>
ورودی	۱/۷۷	۰/۰۱	۳/۹۱	۷/۰۹	Cl <sup>-1</sup>
ورودی	۹/۶۷	۰/۰۰	۰/۰۸	۰/۰۱	CO <sub>3</sub>
ورودی	۱۹/۰۹	۰/۰۰	۲/۴۶	۲/۹۱	HCO <sub>3</sub>
ورودی	۰/۵۱	۰/۰۰	۰/۰۳	۰/۰۵	K <sup>+1</sup>
ورودی	۰/۸۷	۰/۰۰	۰/۹۷	۲/۴۰	Mg <sup>+2</sup>
ورودی	۱/۶۹	۰/۰۱	۳/۹۲	۷/۱۲	Na <sup>+1</sup>
ورودی	۱/۰۷	۰/۰۱	۲/۰۲	۳/۹۹	SO <sub>4</sub>
خروجی	۱/۲۵	۲/۲۲	۵۴۳/۱۰	۱۳۹۲/۷۴	EC
خروجی	۱/۰۲	۱/۴۰	۳۴۳/۴۰	۸۷۸/۷۹	TDS

دیگر استفاده از روش مذکور، کارایی و زمان محاسبه کمتر LSSVM نسبت به SVM (که از روش بهینه‌سازی مرتبه دوم استفاده می‌کند) است (۲۳). همچنین LSSVM تنها از سه لایه تشکیل شده است (یک لایه ورودی، یک لایه پنهان و یک لایه خروجی). بنابراین، مانند ANN نیازی به تعیین تعداد لایه‌های پنهان و نرون‌های هر لایه ندارد. ارتباط رگرسیونی خطی بین ورودی‌ها و خروجی‌ها در الگوریتم LSSVM از معادله ۱، حاصل می‌شود:

$$M = \sum_{i=1}^n K(x, x_i) \alpha_i + b \quad (1)$$

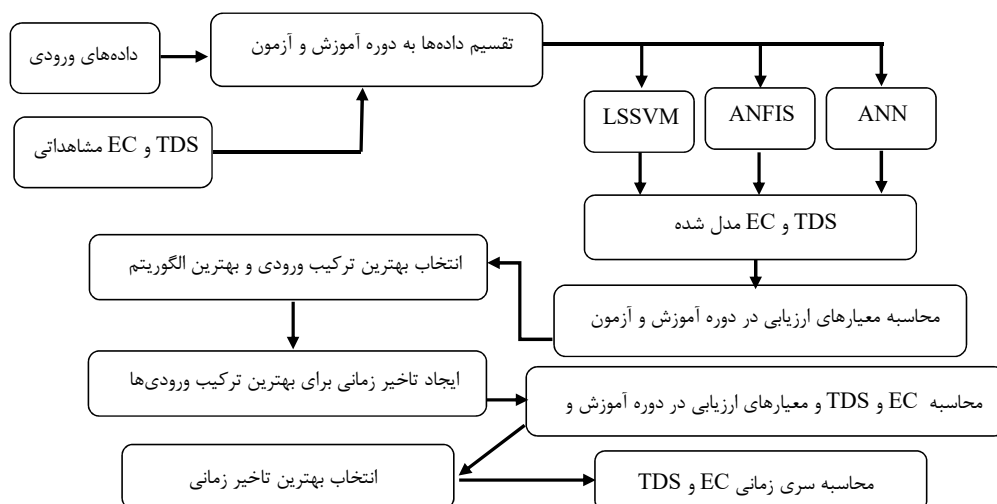
که در آن:  $M$ ،  $\alpha$ ،  $K$  و  $b$  به ترتیب معرف خروجی مدل، ضرایب لاگرانژ، تابع کرنل و بایس مدل هستند. LSSVM به منظور تبدیل رابطه غیرخطی ورودی‌ها و خروجی‌ها به یک رابطه خطی از تابع کرنل استفاده می‌کند. پارامترهای  $\alpha$  و  $b$  نیز با استفاده از حل معادله خطی معادله ۲ محاسبه می‌شوند:

$$\begin{bmatrix} Kernel & \bar{1}^T \\ \bar{1} & Kernel + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ M \end{bmatrix} \quad (2)$$

منظور ابتدا جهت آنالیز حساسیت، شوری آب EC و TDS با استفاده از پارامترهای مختلف آلوده کننده مدل می‌شود. سپس، پارامترهای آلوده کننده‌ای که بیشترین تاثیر را در EC و TDS دارند، انتخاب می‌شود. در ادامه، بهترین الگوریتم (از نظر بیشترین دقت) و موثرترین پارامترهای آلوده کننده جهت مدلسازی با تاخیرهای زمانی متفاوت استفاده می‌شوند و بهترین تاخیر زمانی انتخاب می‌شود. در نهایت، سری زمانی EC و TDS محاسبه می‌شود. شکل ۲، فلوچارت روش مدلسازی شوری آب در پژوهش حاضر را نشان می‌دهد.

### - ماشین بردار پشتیبان حداقل مربعات (LSSVM)

LSSVM اولین بار توسط Suykens (۲۲) به منظور تخمین تابع، طبقه‌بندی و شناسایی الگو به جامع علمی معرفی شد. این الگوریتم با نگاشت ورودی‌ها به فضای با ابعاد بالاتر، رابطه غیرخطی بین ورودی‌ها و خروجی‌ها را به یک رابطه خطی تبدیل می‌کند. این الگوریتم به دلیل استفاده از روش بهینه‌سازی سراسری حداقل مربعات مانند ANN در دام بهینه‌های محلی نمی‌افتد. همچنین مزیت



شکل ۲- روش مدلسازی کیفیت آب



که C معرف ضریب پنالتی است. پارامترهای  $M, \alpha, 1$  و I نیز به شکل زیر هستند (معادله ۳):

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}, \bar{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{bmatrix} \quad (3)$$

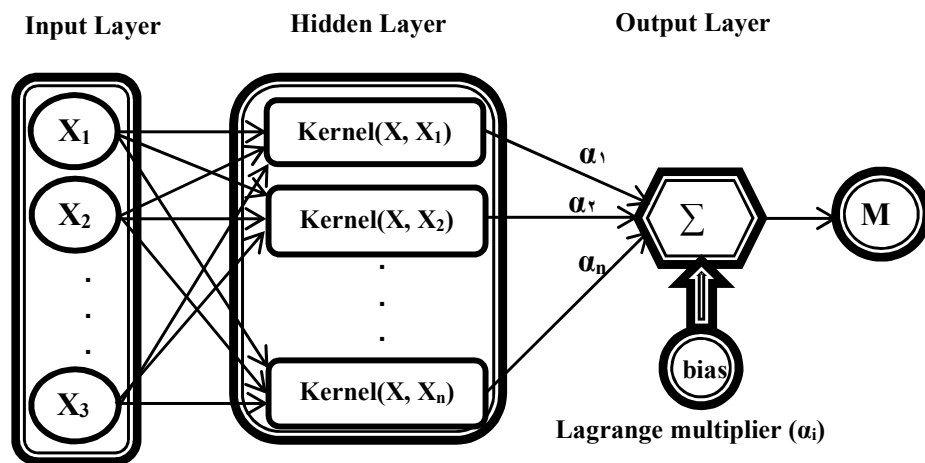
روابط غیرخطی بین ورودی‌ها و خروجی‌ها بهره می‌برد (۲۵). سه روش مختلف شامل تقسیم‌بندی شبکه، خوشه‌بندی براساس مراکز خوشه‌ها و خوشه‌بندی فازی وجود دارد. اما، مطابق با مطالعات انجام شده (۱۹-۲۱) در بحث مدلسازی کیفیت آب روش خوشه‌بندی فازی از دقت بیشتری برخوردار است. بنابراین، در پژوهش حاضر از روش مذکور جهت ایجاد مدل فازی در ANFIS استفاده شده است. شکل ۴، طرح شماتیک روش ANFIS را نشان می‌دهد. مدل ANFIS از پنج لایه تشکیل شده است. لایه اول ورودی‌ها (x,y) و مجموعه‌های فازی (A<sub>1</sub>, A<sub>2</sub>, B<sub>1</sub> and B<sub>2</sub>) هستند.

در پژوهش حاضر مطابق با مطالعه Jang (۲۴) از تابع کرنل شعاعی استفاده شده است. فرمول این تابع به شکل معادله ۴ بیان می‌شود:

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{\sigma^2}\right) \quad (4)$$

شکل ۳ نمودار LSSVM را نشان می‌دهد.

- روش استنتاج تطبیقی فازی-عصبی (ANFIS) -  
الگوریتم ANFIS اولین بار توسط Jang (۲۴) با ترکیب روش فازی و ANN توسعه داده شد. ANFIS از توانایی یادگیری ANN و استنتاج روش فازی به منظور تشخیص



شکل ۳- طرح شماتیک LSSVM

آنها روش LevenbergMarquart است. در پژوهش حاضر نیز از این الگوریتم جهت آموزش ANN استفاده شده است.

#### - معیارهای ارزیابی

در پژوهش حاضر از معیارهای ارزیابی میانگین مطلق خطا MAE، مجذور میانگین مربعات خطا RMSE، مجذور میانگین مربعات خطای نسبی RRMSE، ضریب همبستگی R و ضریب تعیین  $R^2$  جهت ارزیابی نتایج استفاده شده است. مقدار پایین معیارهای MAE، RMSE نشان دهنده عملکرد مطلوب مدل است. کمترین مقدار معیارهای مذکور برابر با صفر است. معیار RRMSE نیز مقدار نرمالایز شده RMSE است. مقدار کمتر از ۰/۵ تا این معیار نشان دهنده عملکرد خیلی خوب، مقدار ۰/۵ تا ۰/۶ نشان دهنده عملکرد خوب و مقدار ۰/۶ تا ۰/۷ نشان دهنده عملکرد متوسط مدل است (۲۶). کمترین مقدار این پارامتر برابر با صفر است (۲۷). مقدار بالای معیارهای R و  $R^2$  نیز نشان دهنده دقت بالای مدل است. بیشترین مقدار این دو معیار برابر با ۱ است. برای معیار  $R^2$  محدوده بین ۰/۶ تا ۰/۷۵ متوسط، ۰/۷۵ تا ۰/۸ خوب و ۰/۸ تا ۱ خیلی خوب است. معیارهای MAE، RMSE و R توسط Ehteram و همکاران (۲۸) مورد استفاده قرار گرفته است. معیارهای RRMSE و  $R^2$  نیز در مطالعات مختلف (۲۷) به کار برده شده‌اند.

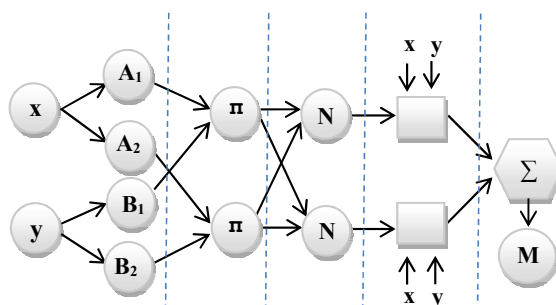
در این لایه به هر ورودی یک تابع عضویت یک عدد فازی اختصاص داده می‌شود. در لایه دوم اعداد فازی لایه اول در یکدیگر ضرب می‌شوند. در ادامه، خروجی‌های لایه دوم در لایه سوم نرمال می‌شوند. در لایه چهارم، هر نرون با استفاده از معادله ۵ یک خروجی ایجاد می‌کند.

$$O_i = \overline{W}_i (p_i x + q_i y + r_i) \quad (5)$$

که در آن:  $p$ ،  $q$  و  $r$  پارامترهای توالی هستند.  $O_i$  و  $\overline{W}_i$  نیز بیانگر خروجی لایه سوم و لایه چهارم هستند. در نهایت نیز، خروجی‌های لایه چهارم با یکدیگر جمع می‌شوند و خروجی نهایی تولید می‌شود (۲۵).

#### - شبکه‌های عصبی مصنوعی (ANN)

ANN یک روش هوشمند، طراحی شده براساس شبکه عصبی مغز انسان است. ساختار این الگوریتم شامل تعدادی نرون در لایه ورودی، یک یا چند لایه پنهان، لایه خروجی، لینک‌های ارتباطی و وزن‌های مربوط به آنها است. تعداد نرون‌های لایه ورودی برابر با تعداد ورودی‌ها و تعداد نرون‌های لایه پنهان نیز با استفاده از سعی و خطا مشخص می‌شود. روش شبکه‌های عصبی با تنظیم وزن‌های مربوط به لینک‌ها رابطه غیرخطی بین ورودی‌ها و خروجی‌ها را تخمین می‌زند. تاکنون روش‌های مختلفی برای تعیین این وزن‌ها تعیین شده است که معروف‌ترین



شکل ۴- طرح شمانیک روش ANFIS



## یافته‌ها

### - آنالیز همبستگی

در پژوهش حاضر معماری ANN براساس سعی و خطا تعیین شده است. بر این اساس تعداد لایه‌ها و نرون‌های لایه مخفی به ترتیب برابر با یک لایه و ۵ نرون در نظر گرفته شده است. در این بخش از آزمون همبستگی پیرسون (Correlation Pierson Test) و نرم‌افزار Matlab 2016 استفاده شده است. بیشترین همبستگی پارامترهای EC و TDS با پارامترهای ورودی، مربوط به  $Na^{+1}$  و  $Cl^{-1}$  (در سطح معنی‌داری ۱ درصد) می‌شود. در هر سه ایستگاه، همبستگی بین EC با ورودی‌ها بیشتر از همبستگی بین

TDS با ورودی‌ها است. همچنین، پارامتر EC در ایستگاه اهواز و پارامتر TDS در ایستگاه ملاثانی دارای بیشترین همبستگی با ورودی‌ها هستند. این همبستگی بیشتر برای برخی پارامترها می‌تواند موجب مدلسازی بهتر این پارامترها شود. همچنین، همبستگی بیشتر پارامترهای EC و TDS در ایستگاه‌های اهواز و ملاثانی می‌تواند دقت بیشتر مدلسازی در این ایستگاه‌ها را به دنبال داشته باشد. در جدول ۲ نتایج آنالیز حساسیت به همراه مقادیر p برای ایستگاه اهواز به صورت نمونه آورده شده است. مقادیر p کوچک‌تر از ۰/۰۱ نشان دهنده همبستگی با سطح معنی‌داری ۰/۰۱ است.

جدول ۲- همبستگی بین پارامترهای کیفیت آب در ایستگاه اهواز

همبستگی											معیار	
TDS	EC	SO <sub>4</sub>	Na <sup>+1</sup>	Mg <sup>+2</sup>	K <sup>+1</sup>	HC	CO <sub>3</sub>	Cl <sup>-1</sup>	Ca <sup>+2</sup>	Q		
										۱/۰۰	R	Q
										۰/۰۰	p	
									۱/۰۰	-۰/۰۳	R	Ca <sup>+2</sup>
									۰/۰۰	۰/۴۶	p	
								۱/۰۰	۰/۶۰	-۰/۴۴	R	Cl <sup>-1</sup>
								۰/۰۰	۰/۰۰	۰/۰۰	p	
							۱/۰۰	۰/۱۲	-۰/۰۱	-۰/۰۷	R	CO <sub>3</sub>
							۰/۰۰	۰/۷۹	۰/۱۳	۰/۱۳	p	
						۱/۰۰	-۰/۱۲	۰/۲۸	۰/۲۳	-۰/۲۵	R	HCO <sub>3</sub>
						۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	
					۱/۰۰	۰/۱	۱/۰۰	۰/۴۴	-۰/۴۸	-۰/۰۶	R	K <sup>+1</sup>
					۰/۰۰	۰/۰۲	۰/۰۰	۰/۰۰	۰/۰۰	۰/۱۵	p	
				۱/۰۰	۰/۳۸	۰/۳۲	۰/۰۸	۰/۷۶	۰/۳۷	-۰/۴۶	R	Mg <sup>+2</sup>
				۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	
			۱/۰۰	۰/۷۶	۰/۴۶	۰/۲۹	۰/۱۳	۰/۹۹	۰/۵۹	-۰/۴۴	R	Na <sup>+1</sup>
			۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	
		۱/۰۰	۰/۷۷	۰/۶۹	۰/۵۳	۰/۱۱	۰/۰۶	۰/۷۴	۰/۸۴	-۰/۲۱	R	SO <sub>4</sub>
		۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۱	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	
	۱/۰۰	۰/۸۵	۰/۹۷	۰/۷۸	۰/۵۰	۰/۳۰	۰/۱۲	۰/۹۷	۰/۷۲	-۰/۴۰	R	EC
	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	
۱/۰۰	۰/۹۴	۰/۸۲	۰/۹۰	۰/۷۲	۰/۴۸	۰/۲۶	۰/۱۳	۰/۸۹	۰/۷۱	-۰/۳۵	R	TDS
۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	p	

- آنالیز حساسیت و الگوریتم منتخب

معیارهای ارزیابی دقت در مدلسازی پارامترهای کیفی EC و TDS ایستگاه اهواز (به‌عنوان نمونه) توسط الگوریتم‌های مختلف به ترتیب در جدول ۳ و ۴ ارائه شده است. با توجه به نتایج مدلسازی EC در جدول ۳، در هر دو دوره

آموزش و آزمون و برای هر سه الگوریتم ترکیب ورودی Q و  $Na^{+1}$  دارای بهترین نتایج است. همچنین، الگوریتم LSSVM در میان سه الگوریتم با مقادیر ۱۳۰/۳۶ برای MAE، ۰/۲۵ برای RRMSE و ۰/۹۷ برای R از بهترین نتایج برخوردار است.

جدول ۳- معیارهای ارزیابی دقت در مدلسازی پارامترهای کیفی EC در ایستگاه اهواز توسط الگوریتم‌های مختلف

پارامتر								معیار	الگوریتم
Q, SO <sub>4</sub>	Q, Na <sup>+1</sup>	Q, Mg <sup>+2</sup>	Q, K <sup>+1</sup>	Q, HCO <sub>3</sub>	Q, CO <sub>3</sub>	Q, Cl <sup>-1</sup>	Q, Ca <sup>+2</sup>		
۱۴۹/۱۴	۸۷/۴۳	۲۱۷/۸۳	۲۸۹/۳۰	۲۸۹/۹۷	۳۰۵/۵۹	۸۵/۳۳	۱۵۷/۶۱	MAE	ANN_Train
۰/۴۶	۰/۲۹	۰/۶۶	۰/۸۶	۰/۸۵	۰/۸۸	۰/۲۷	۰/۵۰	RRMSE	ANN_Train
۰/۷۹	۰/۹۲	۰/۵۶	۰/۳۸	۰/۲۸	۰/۲۶	۰/۹۲	۰/۷۶	R	ANN_Train
۱۳۵/۷۸	۸۱/۴۹	۲۰۳/۲۰	۲۲۵/۷۵	۲۶۴/۷۶	۲۷۸/۶۱	۸۵/۱۳	۱۶۴/۴۸	MAE	ANFIS_Train
۰/۴۲	۰/۲۸	۰/۶۳	۰/۷۰	۰/۷۸	۰/۸۱	۰/۲۸	۰/۵۱	RRMSE	ANFIS_Train
۰/۸۳	۰/۹۲	۰/۶۱	۰/۵۲	۰/۴۰	۰/۳۴	۰/۹۲	۰/۷۴	R	ANFIS_Train
۱۳۱/۲۱	۸۷/۶۳	۱۹۵/۶۱	۲۳۱/۹۹	۲۵۲/۴۵	۲۶۷/۰۵	۹۰/۱۸	۱۷۵/۲۲	MAE	LSSVM_Train
۰/۴۲	۰/۳۰	۰/۶۰	۰/۷۴	۰/۷۶	۰/۸۰	۰/۲۹	۰/۵۶	RRMSE	LSSVM_Train
۰/۸۳	۰/۹۲	۰/۶۴	۰/۴۶	۰/۴۲	۰/۳۶	۰/۹۲	۰/۶۹	R	LSSVM_Train
۲۳۵/۴۹	۱۰۵/۶۲	۳۳۰/۷۹	۲۹۲/۰۲	۴۷۲/۵۵	۵۱۴/۷۰	۱۹۰۰/۷۰	۲۰۶/۷۵	MAE	ANN_Test
۰/۵۱	۰/۳۲	۰/۶۹	۰/۵۹	۰/۹۷	۱/۰۳	۰/۶۹	۰/۴۷	RRMSE	ANN_Test
۰/۷۹	۰/۹۲	۰/۶۴	۰/۶۷	۰/۴۲	۰/۲۸	۰/۷۲	۰/۷۹	R	ANN_Test
۲۰۵/۸۱	۱۲۹/۴۳	۳۰۵/۲۲	۳۴۲/۵۲	۴۰۷/۶۱	۴۸۰/۱۹	۱۳۹/۰۳	۲۱۶/۰۹	MAE	ANFIS_Test
۰/۴۴	۰/۲۶	۰/۶۵	۰/۶۹	۰/۸۴	۰/۹۵	۰/۲۷	۰/۴۷	RRMSE	ANFIS_Test
۰/۸۱	۰/۹۴	۰/۶۴	۰/۶۴	۰/۵۸	۰/۴۶	۰/۹۴	۰/۷۹	R	ANFIS_Test
۲۲۵/۲۰	۱۳۰/۳۶	۳۱۰/۹۹	۳۴۱/۳۳	۴۶۲/۵۶	۵۱۲/۷۱	۱۳۷/۸۶	۲۳۸/۰۱	MAE	LSSVM_Test
۰/۴۹	۰/۲۵	۰/۶۵	۰/۷۰	۰/۹۵	۱/۰۳	۰/۲۶	۰/۴۹	RRMSE	LSSVM_Test
۰/۷۹	۰/۹۴	۰/۶۴	۰/۶۶	۰/۳۸	۰/۳۵	۰/۹۶	۰/۸۱	R	LSSVM_Test

جدول ۴- معیارهای ارزیابی دقت در مدلسازی پارامترهای کیفی TDS در ایستگاه اهواز توسط الگوریتم‌های مختلف

پارامتر								معیار	الگوریتم
Q, SO <sub>4</sub>	Q, Na <sup>+1</sup>	Q, Mg <sup>+2</sup>	Q, K <sup>+1</sup>	Q, HCO <sub>3</sub>	Q, CO <sub>3</sub>	Q, Cl <sup>-1</sup>	Q, Ca <sup>+2</sup>		
۹۷/۶۸	۷۰/۰۶	۱۴۹/۲۳	۱۶۹/۴۴	۲۰۱/۹۳	۱۸۷/۲۹	۸۳/۱۹	۱۲۴/۲۷	MAE	ANN_Train
۰/۴۷	۰/۳۴	۰/۶۸	۰/۷۷	۰/۸۶	۰/۸۱	۰/۳۷	۰/۵۹	RRMSE	ANN_Train
۰/۷۷	۰/۸۸	۰/۵۳	۰/۴۱	۰/۲۹	۰/۳۵	۰/۸۸	۰/۶۶	R	ANN_Train
۹۷/۰۷	۶۸/۶۱	۱۴۵/۶۱	۱۵۴/۴۳	۱۸۸/۰۸	۱۹۵/۷۱	۷۰/۴۷	۱۱۶/۴۷	MAE	ANFIS_Trai

ادامه جدول ۴- معیارهای ارزیابی دقت در مدلسازی پارامترهای کیفی TDS در ایستگاه اهواز توسط الگوریتم‌های مختلف

پارامتر								معیار	الگوریتم
Q, SO <sub>4</sub>	Q, Na <sup>+1</sup>	Q, Mg <sup>+2</sup>	Q, K <sup>+1</sup>	Q, HCO <sub>3</sub>	Q, CO <sub>3</sub>	Q, Cl <sup>-1</sup>	Q, Ca <sup>+2</sup>		
۰/۴۶	۰/۳۴	۰/۶۷	۰/۷۲	۰/۸۳	۰/۸۶	۰/۳۴	۰/۵۵	RRMSE	ANFIS_Trai
۰/۷۹	۰/۸۸	۰/۵۶	۰/۴۸	۰/۳۱	۰/۲۶	۰/۸۸	۰/۶۹	R	ANFIS_Trai
۹۷/۲۰	۷۰/۳۳	۱۴۷/۴۳	۱۵۸/۳۴	۱۷۹/۱۱	۱۸۳/۵۳	۷۷/۶۱	۱۲۳/۴۳	MAE	LSSVM_Tra
۰/۴۶	۰/۳۵	۰/۶۷	۰/۷۳	۰/۸۰	۰/۸۱	۰/۳۵	۰/۵۷	RRMSE	LSSVM_Tra
۰/۷۹	۰/۸۸	۰/۵۵	۰/۴۶	۰/۳۷	۰/۳۴	۰/۸۸	۰/۶۷	R	LSSVM_Tra
۱۷۲/۵۶	۱۲۱/۴۴	۲۳۸/۶۴	۲۴۸/۷۵	۲۹۲/۵۲	۳۰۵/۱۱	۱۲۴/۰۳	۱۸۲/۸۸	MAE	ANN_Test
۰/۶۳	۰/۵۳	۰/۷۸	۰/۸۰	۰/۹۴	۰/۹۵	۰/۵۶	۰/۶۵	RRMSE	ANN_Test
۰/۶۲	۰/۷۲	۰/۴۴	۰/۴۹	۰/۲۹	۰/۳۸	۰/۶۹	۰/۶۱	R	ANN_Test
۱۶۶/۱۹	۱۱۵/۵۵	۲۳۷/۳۴	۲۵۲/۹۸	۳۱۸/۴۶	۳۶۵/۲۰	۱۲۶/۶۸	۱۶۳/۱۴	MAE	ANFIS_Test
۰/۶۲	۰/۵۴	۰/۸۶	۰/۸۵	۱/۰۲	۱/۰۸	۰/۵۵	۰/۶۲	RRMSE	ANFIS_Test
۰/۶۲	۰/۷۱	۰/۳۵	۰/۳۷	۰/۲۲	۰/۲۵	۰/۷۱	۰/۶۱	R	ANFIS_Test
۱۶۳/۵۱	۱۱۸/۳۵	۲۲۲/۷۷	۲۳۶/۶۵	۳۱۶/۰۱	۳۴۳/۸۸	۱۲۶/۷۸	۱۶۳/۱۶	MAE	LSSVM_Tes
۰/۶۱	۰/۵۳	۰/۷۶	۰/۷۸	۰/۹۸	۱/۰۳	۰/۵۴	۰/۶۰	RRMSE	LSSVM_Tes
۰/۶۲	۰/۷۲	۰/۴۸	۰/۴۸	۰/۲۲	۰/۲۶	۰/۷۲	۰/۶۶	R	LSSVM_Tes

سطح آب رودخانه کارون می‌تواند سبب این امر شود. همچنین، مقایسه ضریب تعیین محاسبه شده برای الگوریتم LSSVM (دقیق‌ترین الگوریتم) و پارامتر EC در ایستگاه‌های ملاتانی، اهواز و فارسیات نشان می‌دهد که پارامتر EC در دوره آموزش در هر سه ایستگاه با دقت بالایی مدلسازی شده است. اما، در دوره آزمون بیشترین کمترین دقت به ترتیب مربوط به ایستگاه اهواز و ملاتانی می‌شود. به طوری که ضریب تعیین برای ایستگاه اهواز نسبت به ایستگاه ملاتانی و فارسیات به ترتیب ۱۶ و ۱ درصد بیشتر است. همچنین، مقدار ضریب تعیین برای هر سه ایستگاه در دوره آزمون در محدوده ۰/۸۲ تا ۰/۹۵ قرار دارد.

مقدار ضریب تعیین برای مدلسازی TDS با الگوریتم LSSVM (دقیق‌ترین الگوریتم) در سه ایستگاه بررسی شده، نشان می‌دهد که مدلسازی TDS در دوره آموزش و هر سه ایستگاه با دقت خوبی انجام شده است. اما، در دوره آزمون، ایستگاه ملاتانی دارای بیشترین ضریب تعیین

نتایج محاسبه معیارهای ارزیابی برای مدلسازی TDS در جدول ۴ نشان می‌دهد که ورودی Q و Na<sup>+1</sup> موجب بیشترین دقت در مدلسازی TDS با استفاده از هر سه الگوریتم می‌شود. همچنین الگوریتم LSSVM براساس بیشترین دقت در رتبه اول قرار می‌گیرد. الگوریتم‌های ANFIS و ANN نیز به ترتیب در رتبه‌های دوم و سوم قرار می‌گیرند.

ورودی‌های Na<sup>+1</sup> و Cl<sup>-1</sup> در مدلسازی پارامترهای EC و TDS دارای دقیق‌ترین نتایج هستند. به طوری که مقدار معیار RRMSE برای ورودی‌های مذکور در هر سه ایستگاه در محدوده خیلی خوب قرار دارد. این امر به علت همبستگی بیشتر این ورودی‌ها با پارامترهای مذکور است که در تطابق کامل با نتایج جدول ۲ است. همچنین، تاثیر بیشتر پارامترهای Na<sup>+1</sup> و Cl<sup>-1</sup> بر پارامترهای EC و TDS سه ایستگاه بررسی شده می‌تواند به علت تخلیه پساب حاصل از کشت نیشکر در بالادست این سه ایستگاه باشد. علاوه بر این، ورود آب دریا به رودخانه در اثر پایین آمدن

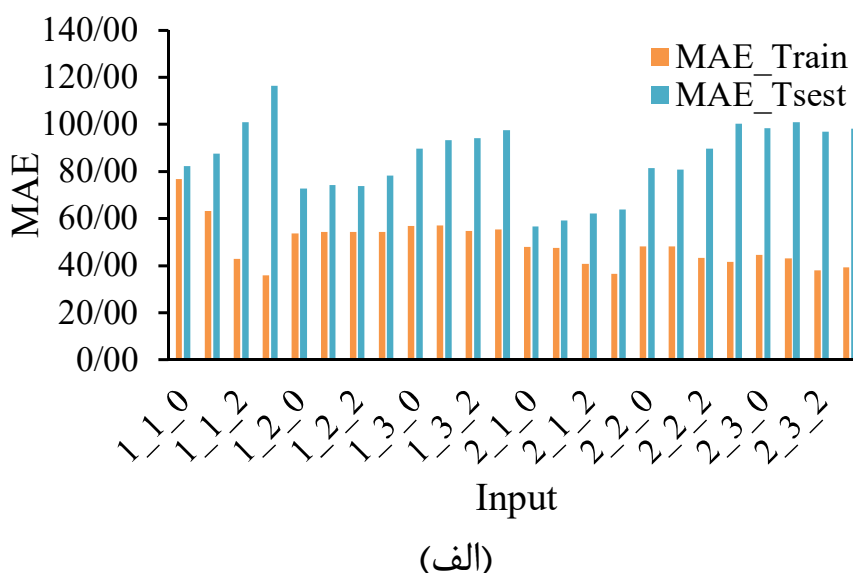
است. با توجه به نمودار ۱، برای پارامتر EC در هر سه ایستگاه بهترین نتایج مربوط به تاخیر زمانی صفر ماه است. برای پارامتر TDS نیز بهترین نتایج برای ایستگاه ملاثانی و فارسیات مربوط به تاخیر زمانی صفر ماه و برای ایستگاه اهواز مربوط به تاخیر زمانی ۲ ماه است. همچنین، بهترین نتایج مدلسازی مربوط به پارامتر EC در ایستگاه اهواز با تاخیر زمانی صفر ماه است. مقدار معیارهای MAE، RRMSE و  $R^2$  برای ایستگاه مذکور به ترتیب برابر با ۰/۷۲/۸۲، ۰/۱۵ و ۰/۹۸ است.

**- سری زمانی مدل شده و مشاهداتی EC و TDS**  
در شکل ۵، نتایج مدلسازی سری زمانی EC براساس بهترین ترکیب ورودی، بهترین تاخیر زمانی و بهترین الگوریتم (LSSVM) در سه ایستگاه ملاثانی، اهواز و فارسیات مورد مقایسه قرار گرفته است. با توجه به این

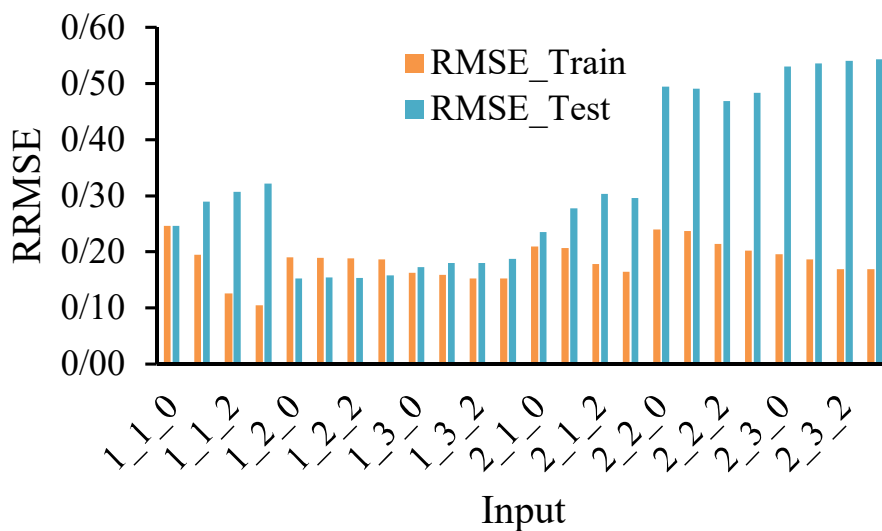
(۰/۹۱) است. مقدار مذکور نسبت به ایستگاه‌های اهواز و فارسیات ۲۶ و ۳۶ درصد بیشتر است. مقدار ضریب تعیین برای ایستگاه ملاثانی در حدود (۰/۹۱) و برای دو ایستگاه دیگر در محدوده (۰/۶۷ تا ۰/۷۷) قرار دارد.

#### - مدلسازی با تاخیرهای زمانی مختلف

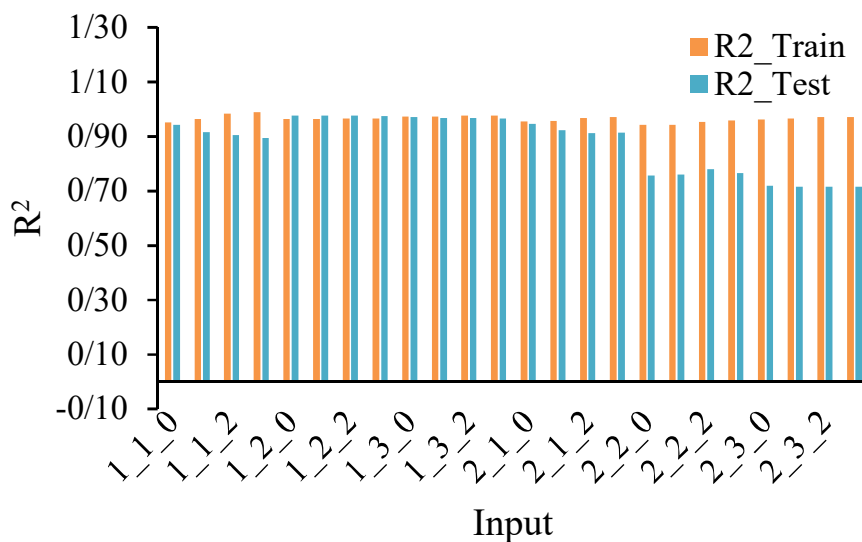
در نمودار ۱ اثر تاخیر زمانی بر نتایج مدلسازی EC و TDS بررسی شده است. به این منظور سه عدد از بهترین پارامترهای کیفی آب با تاخیرهای زمانی ۰ تا ۳ ماه به همراه دبی جریان به عنوان ورودی در نظر گرفته شده‌اند. با توجه به جدول ۳ و ۴ سه پارامتر کیفی شامل  $Na^+$ ،  $Ca^{+2}$  و  $Cl^-$  برای هر سه ایستگاه دارای بهترین نتایج است که برای ادامه مدلسازی از این سه پارامتر استفاده شده است. همچنین از الگوریتم LSSVM که دارای دقیق‌ترین نتایج است جهت مدلسازی EC و TDS استفاده شده



نمودار ۱- معیارهای ارزیابی برای مدلسازی EC و TDS (الف) MAE، (ب) RRMSE، (ج)  $R^2$



(ب)



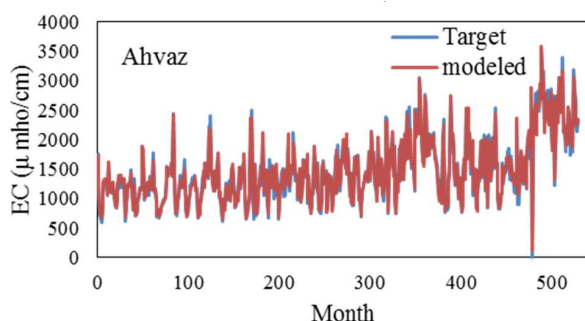
(ج)

ادامه نمودار ۱- معیارهای ارزیابی برای مدل‌سازی EC و TDS الف) MAE، ب) RRMSE، ج) R<sup>2</sup>

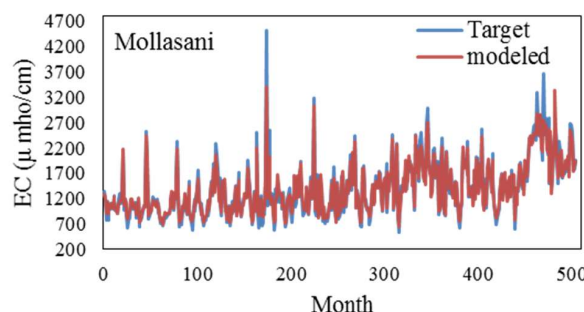
ملاثانی و اهواز با گذشت زمان روند صعودی دارند که این موضوع می‌تواند به دلیل افزایش منابع آلوده کننده باشد. علاوه بر این، مقدار پارامترهای EC و TDS در ایستگاه فارسیات به ترتیب برابر با  $1671/21 \mu\text{mho/cm}$  و  $1065/70 \text{ mg/L}$  است. این در حالی است که مقدار میانگین EC در دو ایستگاه ملاثانی و اهواز در بالادست ایستگاه فارسیات به ترتیب برابر با  $1390/17 \mu\text{mho/cm}$  و  $1475/88 \mu\text{mho/cm}$  است.

شکل مقدار نوسانات EC توسط الگوریتم LSSVM به خوبی مدل شده است که حاکی از دقت بالای این روش و تایید کننده نتایج شکل ۵ است.

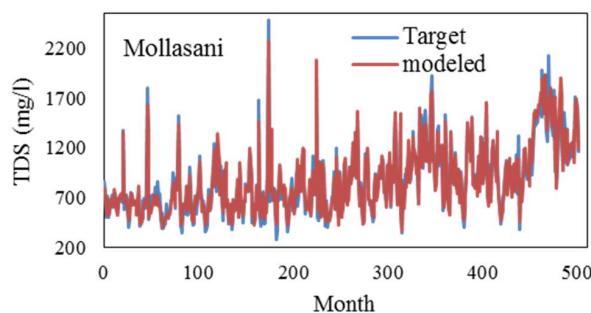
نتایج سری زمانی TDS مدل شده توسط LSSVM نیز در شکل ۵، در سه ایستگاه ملاثانی، اهواز و فارسیات مورد بررسی قرار گرفته است. مطابق با نتایج این شکل، مقدار تغییرات سری زمانی TDS و نقاط پیک آن به خوبی توسط الگوریتم LSSVM مدل شده است. همچنین، براساس شکل ۵ مقدار سری زمانی EC و TDS در دو ایستگاه



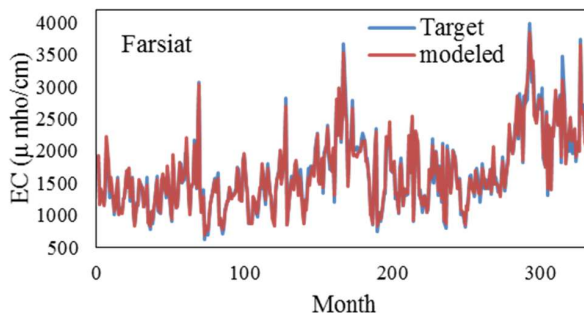
(ب)



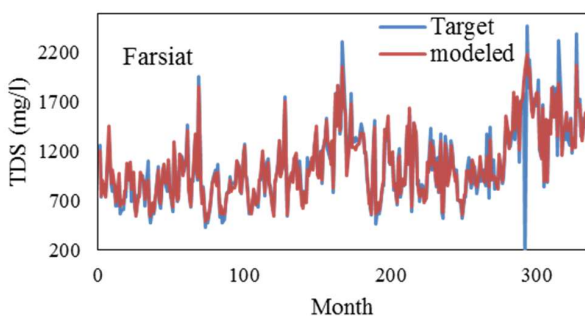
(الف)



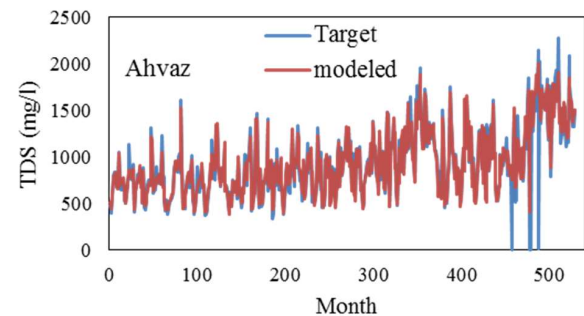
(د)



(ج)



(و)



(ه)

شکل ۵- مقدار EC و TDS مشاهداتی و شبیه‌سازی شده با استفاده از بهترین الگوریتم و ترکیب ورودی الف) EC در ملاثانی، ب) EC در اهواز، ج) EC در فارسیات، د) TDS در ملاثانی، و) TDS در اهواز و ه) TDS در فارسیات



ANN برای مدلسازی جامع EC و TDS مورد استفاده قرار گرفته‌اند. به این منظور EC و TDS توسط هر یک از الگوریتم‌ها براساس ورودی‌های مختلف (پارامترهای آلوده کننده و دبی) در سه ایستگاه ملاتانی، اهواز و فارسیات مدل شده‌اند. نتایج جدول ۳ و ۴ نشان می‌دهد که LSSVM نسبت به دو الگوریتم ANFIS و ANN دارای برتری است. این موضوع تایید کننده نتایج پژوهش Farzin و همکاران (۲۹) است. دلیل برتری LSSVM استفاده از تابع کرنل و تبدیل روابط غیرخطی بین ورودی‌ها و خروجی‌ها به یک رابطه خطی است. همچنین الگوریتم LSSVM در ساختار خود از الگوریتم بهینه‌سازی سراسری حداقل مربعات استفاده می‌کند در حالی که الگوریتم‌های ANFIS و ANN در ساختار خود از الگوریتم‌های بهینه‌سازی محلی استفاده می‌کنند. این موضوع سبب می‌شود که دو الگوریتم مذکور در دام بهینه‌های محلی بیافتند و در نتیجه نسبت به الگوریتم LSSVM عملکرد ضعیف‌تری داشته باشند. برتری ANFIS نسبت به ANN نیز به دلیل ترکیب قدرت یادگیری الگوریتم ANN و توانایی استدلال روش فازی در الگوریتم مذکور است (۲۴).

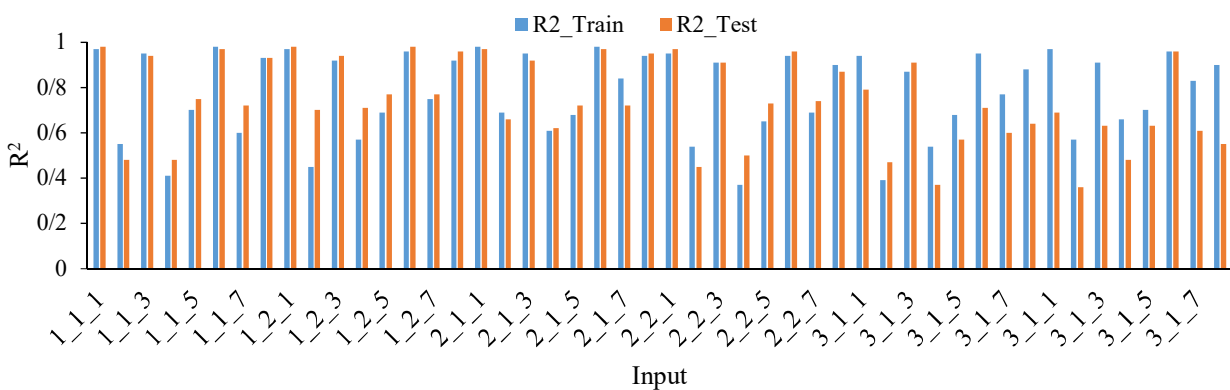
### - مدلسازی با کالیبراسیون تصادفی

در نهایت، به منظور تضمین دقت نتایج مدلسازی، از شبیه‌سازی به روش کالیبراسیون تصادفی استفاده شده است. بدین منظور، ورودی‌های الگوریتم (LSSVM) به صورت تصادفی انتخاب شده‌اند. علت انتخاب الگوریتم مذکور این است که در قسمت قبل، بهترین عملکرد را داشته است. نمودار ۲، نتایج مدلسازی با استفاده از کالیبراسیون تصادفی را نشان می‌دهد. مطابق با نتایج نمودار ۲، الگوریتم LSSVM به خوبی آموزش دیده است. همچنین در دوره آزمون بهترین نتایج مدلسازی، مربوط به پارامترهای  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $Na^{+1}$  است. مقدار ضریب تعیین محاسبه شده برای EC و TDS با استفاده از ورودی‌های مذکور و کالیبراسیون غیرتصادفی به ترتیب در محدود ۰/۶۷ تا ۰/۹۵، ۰/۶۶ تا ۰/۹۶ و ۰/۵۹ تا ۰/۸۶ قرار دارد. مقادیر مذکور برای کالیبراسیون تصادفی در محدود ۰/۶۹ تا ۰/۹۸، ۰/۶۳ تا ۰/۷۷ و ۰/۷۱ تا ۰/۹۷ است.

### بحث

#### - الگوریتم‌های بررسی شده

در پژوهش حاضر الگوریتم‌های LSSVM، ANFIS و



نمودار ۲- نتایج مدلسازی LSSVM با استفاده از کالیبراسیون تصادفی

در این نمودار a\_b\_c به معنی ایستگاه a (۱ برای ملاتانی، ۲ برای اهواز و ۳ برای فارسیات)، پارامتر b (۱ برای EC و ۲ برای TDS) و ورودی c (۱ تا ۸ به ترتیب برای پارامترهای  $Ca^{+2}$ ،  $Cl^{-1}$ ،  $Na^{+1}$ ،  $Mg^{+2}$ ،  $K^{+1}$ ،  $CO_3$ ،  $HCO_3$ ،  $SO_4$ ) است.

**- آنالیز حساسیت پارامترها**

در پژوهش حاضر به منظور تعیین پارامترهایی که بیشترین تاثیر را در مدلسازی EC و TDS دارند از سه روش LSSVM، ANFIS و ANN با ورودی‌های مختلف استفاده شد. نتایج جدول ۳ و ۴ نشان داد، ورودی‌های  $Na^{+1}$ ،  $Ca^{+2}$  و  $Cl^{-1}$  در مدلسازی پارامترهای EC و TDS دارای دقیق‌ترین نتایج هستند. این امر به علت همبستگی بیشتر این ورودی‌ها با پارامترهای مذکور است که در تطابق کامل با نتایج جدول ۲ است. همچنین، تاثیر بیشتر پارامترهای  $Na^{+1}$ ،  $Ca^{+2}$  و  $Cl^{-1}$  بر پارامترهای EC و TDS سه ایستگاه بررسی شده می‌تواند به علت ماهیت شیمیایی آنها باشد که سبب افزایش هدایت الکتریکی آب و املاح محلول در آب می‌شود. علت دیگر این موضوع، تخلیه پساب حاصل از کشت نیشکر در بالادست این سه ایستگاه است. علاوه بر این، ورود آب دریا به رودخانه در اثر پایین آمدن سطح آب رودخانه کارون می‌تواند سبب این امر شود. دلیل دیگر این موضوع نیز می‌تواند به دلیل ساختار زمین‌شناسی منطقه و ورود املاحی مانند  $Na^{+1}$  و  $Ca^{+2}$  به رودخانه باشد. با این وجود در مورد پارامتر  $Cl^{-1}$  مهمترین عامل تخلیه پساب به رودخانه هست. تاثیر بیشتر پارامترهای  $Na^{+1}$ ،  $Ca^{+2}$  و  $Cl^{-1}$  بر EC و TDS تایید کننده نتایج Azad و همکاران (۲۰) است. در این پژوهش که مطالعه موردی آن رودخانه گرگان رود بود، پارامترهای آلوده کننده مذکور بیشترین همبستگی را با EC و TDS داشتند.

با ورودی‌ها با دقت بیشتری نسبت به TDS مدلسازی شده است. موارد ذکر شده تاییدکننده نتایج جدول ۳ و ۴ هستند. در پژوهش Banejad و همکاران (۱۰) نیز پارامتر EC با دقت بیشتری نسبت به TDS مدل شد. براساس شکل ۵، مقدار EC در پایین‌دست (ایستگاه فارسیات) نسبت به بالادست در حدود ۲۰ درصد افزایش داشته است. مقدار میانگین TDS نیز در دو ایستگاه ملاثانی و اهواز برابر با  $881/88$  و  $931/10$  mg/L است. این مقدار در ایستگاه فارسیات در حدود ۲۱ درصد بیشتر از بالادست (دو ایستگاه ملاثانی و اهواز) است. افزایش EC و TDS در پایین دست رودخانه نسبت به بالادست رودخانه، نشان‌دهنده ورودی پساب‌های آلوده‌کننده به رودخانه کارون در این بازه از رودخانه است.

روش‌های هوشمند مانند مدل‌های فیزیکی و ریاضی نیازی به تعداد زیاد پارامترهای اولیه، شرایط مرزی و تلاش محاسباتی بالا ندارند. این روش‌ها با صرف زمان محاسباتی کمتر و در عین حال دقت قابل قبول به مدلسازی پارامترهای مختلف از جمله کیفیت آب می‌پردازند. همچنین روش‌های هوشمند قابلیت ارزیابی اثر چندین پارامتر آلوده کننده را بر یک پارامتر مانند کیفیت آب دارند. روش‌های داده‌کاوی، ابزاری مفید جهت ارزیابی کیفیت آب رودخانه‌ها و برخلاف روش‌های آماری قابلیت بررسی چندین پارامتر آلوده کننده بر یک پارامتر را دارند (۱۱).

**-مدلسازی با کالیبراسیون تصادفی**

با بررسی نتایج می‌توان دریافت که شبیه‌سازی با استفاده از ترکیب ورودی مربوط به  $Cl^{-1}$  با دقت بیشتری توسط کالیبراسیون غیرتصادفی انجام شده است. در مورد ترکیب‌های ورودی مربوط به  $Na^{+1}$  و  $Ca^{+2}$  دقت شبیه‌سازی با کالیبراسیون تصادفی اندکی بیشتر است. بنابراین، نزدیکی نتایج دو روش متضمن دقت نتایج روش غیرتصادفی است. این موضوع نشان می‌دهد که الگوریتم‌های داده‌کاوی به خوبی کالیبره شدند و طول دوره

**- سری زمانی مدل شده و مشاهداتی EC و TDS**

همچنین به منظور مدلسازی EC و TDS از الگوریتم منتخب (از نظر بیشترین دقت) و موثرترین پارامترهای تعیین شده در بخش آنالیز حساسیت پارامترها، استفاده شده است. دقت بیشتر در مدلسازی EC در ایستگاه اهواز و مدلسازی TDS در ایستگاه ملاثانی به علت همبستگی بیشتر ورودی‌ها در ایستگاه اهواز و ملاثانی با EC و TDS است. همچنین، پارامتر EC به دلیل همبستگی بیشتر

رودخانه کارون توسط روش‌های کالیبراسیون غیرتصادفی و تصادفی، حاکی از عملکرد مناسب هر دو روش بود. بنابراین، می‌توان بیان داشت که روش‌های داده‌کاوی به‌ویژه LSSVM از پتانسیل بالایی برای مدل‌سازی پارامترهای کیفی رودخانه‌ها مانند شوری آب برخوردارند. علاوه بر این می‌توان این مدل‌ها را جهت مدیریت منابع آب و توسعه مناطق نزدیک رودخانه‌ها به‌کار برد. در ادامه پژوهش حاضر و پیشنهاد می‌شود که عدم قطعیت روش‌های بررسی شده در مدل‌سازی EC و TDS مورد ارزیابی قرار بگیرد و حدود اطمینان در مدل‌سازی پارامترهای مذکور تعیین شود.

### ملاحظات اخلاقی

نویسندگان کلیه نکات اخلاقی شامل عدم سرقت ادبی، انتشار دوگانه، تحریف داده‌ها و داده‌سازی را در این مقاله رعایت کرده‌اند.

### تشکر و قدردانی

بدینوسیله مولفین این مقاله تحقیقاتی از همکاری شرکت آب منطقه‌ای استان خوزستان تقدیر و تشکر می‌نمایند.

## References

1. Orouji H, Bozorg Haddad O, Fallah-Mehdipour E, Mariño M. Modeling of water quality parameters using data-driven models. *Journal of Environmental Engineering*. 2013;139(7):947-57.
2. Ravansalar M, Rajaei T. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environmental Monitoring and Assessment*. 2015;187(6):366.
3. Barzegar R, Moghaddam AA, Adamowski J, Ozga-Zielinski B. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stochastic Environmental Research and Risk Assessment*. 2018;32(3):799-813.
4. Haghghi AH, Nasrolahi AH, Parsaie A. Water quality prediction using machine learning methods. *Water Quality Research Journal*. 2018;53(1):3-13.

در نظر گرفته شده جهت آموزش الگوریتم‌های داده‌کاوی کافی بوده و نیازی به تغییر دوره آموزش و یا استفاده از روش‌های دیگر کالیبراسیون نیست.

### نتیجه‌گیری

مدلسازی و پیش‌بینی پارامترهای مربوط به آلودگی رودخانه‌ها، نقش بسزایی در مدیریت منابع آب، سلامت انسان‌ها و حفاظت از محیط زیست ایفا می‌کند. در پژوهش حاضر، از سه الگوریتم LSSVM، ANFIS و ANN جهت مدل‌سازی کیفیت آب (EC و TDS) در سه ایستگاه ملاثانی، اهواز و فارسیات واقع در پایین دست حوضه کارون استفاده شد. به این منظور، پارامترهای کیفی آب شامل،  $SO_4$ ،  $HCO_3$ ،  $CO_3$ ،  $K^{+1}$ ،  $Mg^{+2}$ ،  $Na^{+1}$ ،  $Ca^{+2}$ ،  $Cl^{-1}$  دبی جریان (Q) به‌کار برده شد. نتایج حاکی از برتری LSSVM و تاثیر بیشتر پارامترهای  $Na^{+1}$ ،  $Cl^{-1}$ ،  $Ca^{+2}$  و Q در مدل‌سازی EC و TDS بود. همچنین، نتایج مدل‌سازی با استفاده از الگوریتم منتخب و موثرترین پارامترها حاکی از دقت بالای مدل‌سازی بود. به‌طوری‌که نقاط پیک و نوسانات سری زمانی EC و TDS به‌خوبی مدل شدند. مقایسه یافته‌های حاصل از مدل‌سازی پارامترهای EC و TDS

5. Montaseri M, Ghavidel SZZ, Sanikhani H. Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques. *Stochastic Environmental Research and Risk Assessment*. 2018;32(8):2253-73.
6. Salarijazi M, Ghorbani K. Improvement of the simple regression model for river'EC estimation. *Arabian Journal of Geosciences*. 2019;12(7):235.
7. Al-Mukhtar M, Al-Yaseen F. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in south of Iraq. *Hydrology*. 2019;6(1):24.
8. Singh B. Prediction of the sodium absorption ratio using data-driven models: a case study in Iran. *Geol-*

- ogy, Ecology, and Landscapes. 2020;4(1):1-10.
9. Rajaei T, Rahimi Benmaran R, Jafari H. Prediction of quality parameters (NO<sub>3</sub>, DO) of Karaj River using ANN, MLR, and Denoising-based combined wavelet-neural network based on Models. *Iranian Journal of Health and Environment*. 2015;7(4):511-30 (in Persian).
  10. Banejad H, Kamali M, Amirmoradi K, Olyaie E. 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 Journal of Health and Environment*. 2013;6(3):277-94 (in Persian).
  11. Soleimanpour S, Mesbah S, Hedayati B. Application of CART decision tree data mining to determine the most effective drinking water quality factors (case study: Kazeroon plain, Fars province). *Iranian Journal of Health and Environment*. 2018;11(1):1-14 (in Persian).
  12. Naddafi K, Honari H, Ahmadi M. Water quality trend analysis for the Karoon River in Iran. *Environmental Monitoring and Assessment*. 2007;134(1-3):305-12.
  13. Musavi-Jahromi S, Golabi M. Application of artificial neural networks in the river water quality modeling: Karoon River, Iran. *Journal of Applied Sciences*. 2008;8(12):2324-28.
  14. Salami E, Salari M, Ehteshami M, Bidokhti N, Ghadimi H. Application of artificial neural networks and mathematical modeling for the prediction of water quality variables (case study: south-west of Iran). *Desalination and Water Treatment*. 2016;57(56):27073-84.
  15. Ehteram M, Mousavi SF, Karami H, Farzin S, Singh VP, Chau K-W, et al. Reservoir operation based on evolutionary algorithms and multi-criteria decision-making under climate change and uncertainty. *Journal of Hydroinformatics*. 2018;20(2):332-55.
  16. Valikhan Anaraki M, Farzin S, Mousavi S-F, Karami H. Application of hybrid least squares support vector machine-whale algorithm (LSSVM\_WOA) for downscaling and prediction precipitation under climate change (A case study: Karun 3 Basin). *Journal of Irrigation and Water Engineering*. 2019. DOI: 10.22125/iwe.2019.174180.1085 (in Persian).
  17. Ghorbani M, Naghipour L, Karimi V, Farhoudi R. Sensitivity Analysis of the Effective Input Parameters upon the Ozone Concentration using Artificial Neural Networks. *Iranian Journal of Health and Environment*. 2013;6(1):11-22 (in Persian).
  18. Valikhan Anaraki M, Mousavi S-F, Farzin S, Karami H. Introducing a nonlinear model based on hybrid machine learning for modeling and prediction of precipitation and comparison with SDSM method (Cases studies: Shahrekord, Barez, and Yasuj). *Iranian Journal of Soil and Water Research*. 2020;51(2):325-39 (in Persian).
  19. Azad A, Farzin S, Kashi H, Sanikhani H, Karami H, Kisi O. Prediction of river flow using hybrid neuro-fuzzy models. *Arabian Journal of Geosciences*. 2018;11(22):718.
  20. Azad A, Karami H, Farzin S, Saeedian A, Kashi H, Sayyahi F. Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (Case study: Gorganrood River). *KSCE Journal of Civil Engineering*. 2018;22(7):2206-13.
  21. Azad A, Manoochehri M, Kashi H, Farzin S, Karami H, Nourani V, et al. Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling. *Journal of Hydrology*. 2019;571:214-24.
  22. Suykens JA. Nonlinear modelling and support vector machines. IMTC 2001. p Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188); 21-23 May 2001; Budapest, Hungary.
  23. Ghosh S. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. *Journal of Geophysical Research: Atmospheres*. 2010;115(D22).
  24. Jang J-S. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*. 1993;23(3):665-85.
  25. Azad A, Karami H, Farzin S, Mousavi S-F, Kisi O. Modeling river water quality parameters using modified adaptive neuro fuzzy inference system. *Water Science and Engineering*. 2019;12(1):45-54.
  26. Babaei M, Moeini R, Ehsanzadeh E. Artificial neu-

- ral network and support vector machine models for inflow prediction of dam reservoir (Case study: Zayandehroud Dam Reservoir). *Water Resources Management*. 2019;33(6):2203-18 (in Persian).
27. Valikhan-Anaraki M, Mousavi S-F, Farzin S, Karami H, Ehteram M, Kisi O, et al. Development of a novel hybrid optimization algorithm for minimizing irrigation deficiencies. *Sustainability*. 2019;11(8):2337.
28. Ehteram M, Karami H, Mousavi S-F, Farzin S, Kisi O. Evaluation of contemporary evolutionary algorithms for optimization in reservoir operation and water supply. *Journal of Water Supply: Research and Technology—AQUA*. 2018;67(1):54-67.
29. Farzin S, Nabizadeh Chianeh F, Valikhan Anaraki M, Mahmoudian F. Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID). *Journal of Cleaner Production*. 2020;266:122075.



Available online: <http://ijhe.tums.ac.ir>

Original Article



## Modeling of qualitative parameters (Electrical conductivity and total dissolved solids) of Karun river at Mollasani, Ahvaz and Farsiat stations using data mining methods

Alireza Mohaghegh, Mahdi Valikhan Anaraki, Saeed Farzin\*

Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran

### ARTICLE INFORMATION:

**Received:** 7 March 2020

**Revised:** 30 May 2020

**Accepted:** 2 June 2020

**Published:** 20 June 2020

**Keywords:** Water salinity, Data mining, Karun basin, Random calibration, Non-random calibration

**\*Corresponding Author:**

saeed.farzin@semnan.ac.ir

### ABSTRACT

**Background and Objective:** In the present study, EC and TDS quality parameters of Karun River were modeled using data-mining algorithms including LSSVM, ANFIS, and ANN, at Mollasani, Ahvaz and Farsiat hydrometric stations.

**Material and Methods:** Eight different inputs including the combination of  $Cl^{-1}$ ,  $Ca^{+2}$ ,  $Na^{+1}$ ,  $Mg^{+2}$ ,  $K^{+1}$ ,  $CO_3^{2-}$ ,  $HCO_3$ , and  $SO_4^{2-}$  with discharge flow (Q) were selected as non-random and random calibration inputs for these algorithms. Then, in order to guarantee the accuracy of the results, the simulation was performed by random calibration and the results of the two methods were compared. In the next step, the EC and TDS parameters were modeled based on the four parameters of  $Na^{+1}$ ,  $Cl^{-1}$ ,  $Ca^{+2}$ , and Q and a lag time of zero to three months.

**Results:** Modeling results indicated that  $Na^{+1}$ ,  $Cl^{-1}$ , and  $Ca^{+2}$  have the highest influence on modeling of EC and TDS parameters. The LSSVM algorithm was the most accurate in modeling EC and TDS parameters. Among the studied stations, the highest precision for EC and TDS modeling belongs to Ahvaz and Mollasani station, which has 16% and 36% higher coefficient of determination. LSSVM has highest accuracy in modeling the oscillation and peak EC and TDS parameters in during times.

**Conclusion:** The methods and models applied in the present study especially the LSSVM algorithm, can be a useful decision-making tool for predicting and qualitative management of rivers, including rivers in the Karun catchment area. The results of modeling the quality parameters of the rivers were reliable and usable by using both non-random and random calibration methods. However, the accuracy of the random calibration method was slightly higher.