

تعیین طبقه‌بندی کیفی آب بر اساس حداقل پارامترهای کیفی (مطالعه موردی: رودخانه کارون)

دنیا دزفولی^۱، سید محمد حسینی موغاری^۲، کیومرث ابراهیمی^{۳*}، شهاب عراقی نژاد^۴

۱. دانش آموخته کارشناسی ارشد مهندسی منابع آب، گروه آبیاری و آبادانی، دانشگاه تهران، کرج، ایران

۲. دانشجوی دکتری مهندسی منابع آب، گروه آبیاری و آبادانی، دانشگاه تهران، کرج، ایران

۳. استاد گروه آبیاری و آبادانی، دانشگاه تهران، کرج، ایران

۴. دانشیار گروه آبیاری و آبادانی، دانشگاه تهران، کرج، ایران

(تاریخ دریافت: ۱۳۹۵/۹/۲۰؛ تاریخ تصویب: ۱۳۹۶/۷/۲۹)

چکیده

رودخانه‌ها یکی از مهم‌ترین منابع تأمین آب شیرین به‌شمار می‌روند. محدودیت این منابع ضرورت حفظ کیفیت آن‌ها را نشان می‌دهد. به‌منظور پایش کیفی منابع آب معمولاً از شاخص‌های کیفیت آب استفاده می‌شود. هر کدام از این شاخص بر اساس پارامترهای کیفی مشخصی محاسبه می‌شوند که فرآیند نمونه‌برداری و تعیین مقدار این پارامترها زمان‌بر و پرهزینه است، لذا یافتن روشی دقیق که در آن با حداقل پارامترهای کیفی بتوان طبقه‌بندی کیفیت آب را تعیین کرد بسیار مفید است. در مقاله حاضر از مزایای شبکه عصبی احتمالی (PNN) به‌عنوان یک طبقه‌بندی‌کننده برای تعیین کیفیت آب رودخانه کارون به‌عنوان جایگزینی برای شاخص متداول و پرکاربرد NSFQI استفاده شد. برای این منظور از آمار کیفی ۱۷۲ نمونه استفاده شد به این صورت که پارامترهای کیفی و کلاس‌های کیفیت آب حاصل از شاخص NSFQI به ترتیب به عنوان ورودی مدل و خروجی مدل در نظر گرفته شدند. جهت ارزیابی عملکرد مدل PNN، از معیارهای ارزیابی نرخ خطا، مقدار خطا، دقت و ضریب همبستگی اسپیرمن استفاده شد. نتایج نشان داد که PNN تنها با استفاده از سه پارامتر کیفی کدورت، کلیرم مدفوعی و کل مواد جامد می‌تواند با دقت ۹۴/۳۷٪ و ۹۰/۷۸٪ به ترتیب در مرحله آموزش و آزمایش، طبقه‌بندی کیفی آب را مشخص کند که بیانگر دقت بالای PNN در تعیین طبقه‌بندی کیفی آب می‌باشد.

کلید واژگان: آب سطحی، پارامترهای کیفی آب، شاخص NSFQI، طبقه‌بندی کیفیت آب، مدل‌سازی

۱. مقدمه

رودخانه‌ها یکی از منابع مهم تأمین آب شیرین برای مصارف شرب، کشاورزی، خانگی و آبیان هستند (Rahman *et al.*, 2014). آلودگی آب توسط آلاینده‌های شیمیایی، فیزیکی و بیولوژیکی در تمام دنیا، یکی از چالش‌های جدی مدیریت منابع آب محسوب می‌شود (Rasinezami *et al.*, 2012). رشد روز افزون جمعیت جهان مصارف گوناگون آب را افزایش داده است. این امر نه تنها کمیت منابع آب در دسترس را کاهش داده و با محدودیت‌های بیشتری روبه‌رو کرده است، بلکه به دلیل توسعه شهرنشینی و فعالیت‌های صنعتی و کشاورزی، تغییر و تنزل کیفیت آب‌ها را نیز به دنبال داشته است. از آنجا که ایران از نظر اقلیمی و جغرافیایی در منطقه‌ای خشک و نیمه‌خشک قرار گرفته است، منابع آب در کشور ما با محدودیت کلی رو به رو هستند، لذا با توجه به کمبود آب و توسعه صنعتی در ایران برنامه‌ریزی و رسیدگی جدی به کیفیت آب رودخانه‌های کشور برای مصارف شرب، کشاورزی و محیط زیست از حیاتی‌ترین موضوعاتی است که بایستی توجه بسیاری به آن نمود.

ارزیابی پارامترهای کیفی آب برای برنامه‌ریزی و توسعه بهتر مدیریت منابع آب لازم است (Walker *et al.*, 2015). به منظور ارزیابی کیفیت آب، تکنیک‌های بسیاری معرفی شده‌اند که یکی از متداول‌ترین روش‌های ارزیابی کیفیت آب استفاده از شاخص‌های کیفی آب^۱ (WQI) است. این روش‌های ساده به دور از پیچیدگی‌های ریاضی و آماری می‌توانند شرایط کیفی آب را بیان کنند. در این روش حجم زیاد اطلاعات حاصل از اندازه‌گیری کیفی آب به صورت یک عدد منفرد و بدون بعد تبدیل شده که در مقیاس درجه‌بندی شده‌ای دارای مفهوم و تفسیر کیفی تعریف شده‌ای می‌گردد (Abbasi *et al.*, 2012). اولین شاخص کیفیت آب در سال ۱۹۶۵ توسط هورتن ارائه شد (Horton, 1965). این شاخص با ارائه یک روش ریاضی ساده عصری جدید را در مطالعات کیفیت آب رقم زد. برخی از متداول‌ترین

شاخص‌های کیفیت آب عبارتند از: NSFQWI^۲ (شاخص کیفیت آب سازمان بهداشت ملی ایالات متحده آمریکا)، OWQI^۳ (شاخص کیفیت آب ارگان)، BCWQI^۴ (شاخص کیفیت آب بریتیش کلمبیا) و CCME-WQI^۵ (شاخص کیفیت آب کانادایی). هریک از این روش‌ها ساختار ریاضی ویژه خود را داشته و نحوه رسیدن به ضرایب وزنی و مشخصه‌های کیفی در آن‌ها متفاوت است. در این روش‌ها، برای تعیین طبقه‌بندی کیفی آب وجود تمامی پارامترهای در نظر گرفته شده لازم است که این امر موجب افزایش هزینه‌های نمونه‌برداری و طولانی شدن محاسبات شده است. از این‌رو استفاده از روش‌های یادگیری ماشین یکی از راهکارهای پیش‌رو برای کاهش هزینه‌های نمونه‌برداری و آزمایشگاهی است و این امکان را فراهم می‌سازد تا در مدت زمان کم و با استفاده از تعداد پارامتر کمتری طبقه کیفی آب تعیین شود. روش‌های یادگیری ماشین به منظور طبقه‌بندی یک ارتباط ریاضی بین مجموعه‌ای از متغیرهای توصیفی (به عنوان مثال اندازه‌گیری‌های شیمیایی) و یک متغیر کیفی (به عنوان مثال عضویت در یک دسته‌بندی تعریف شده) را پیدا می‌کنند (Sakizadeh, 2015).

پژوهش‌های زیادی در سطح جهان و ایران در این زمینه انجام شده است که در ادامه به برخی از آن‌ها اشاره می‌شود. Asgari و Kerachian (۲۰۰۷) با استفاده از ماشین‌های بردار پشتیبان^۶ (SVM) پهنه‌بندی کیفی آب رودخانه کارون-دز را بررسی کردند. نتایج تحقیقات ایشان بیانگر آن است که روش پیشنهاد شده می‌تواند اطلاعات با ارزشی برای پشتیبانی از تصمیم‌گیری و کمک به مدیریت کیفی آب ارائه کند. Sattari و همکاران (۲۰۱۴) با استفاده از روش درخت تصمیم به پیش‌بینی

¹ Water Quality Index

² National Sanitation Foundation Water Quality Index

³ Oregon Water Quality Index

⁴ British Columbia Water Quality Index

⁵ Canadian Council of Ministers of Environment Water Quality Index

⁶ Support Vector Machine

بیشترین مقدار خطا ضعیف‌ترین عملکرد را داشته است. Sharma و همکاران (۲۰۱۴) تغییرات در شاخص کیفیت آب رودخانه گنگ را در مکان‌های مختلف الله‌آباد بررسی کردند. نتایج ایشان نشان داد که WQI یک ابزار مفید برای ارزیابی کیفیت آب و پیش‌بینی روند تغییر در کیفیت آب در مکان‌های مختلف رودخانه گنگ است. Sakizadeh (۲۰۱۵) به ارزیابی عملکرد روش‌های طبقه‌بندی در مطالعات کیفیت آب پرداخت. در این تحقیق عملکرد روش‌های طبقه‌بندی تفکیک خطی^۷ و نایو بیز^۸ در نه ایستگاه نمونه‌برداری رودخانه کرج ایران به کار برده شد. نتایج این تحقیق نشان داد که روش تفکیک خطی در مقایسه با روش طبقه‌بندی نایو بیز از عملکرد بهتری در تعیین طبقه‌بندی کیفی آب این رودخانه برخوردار است.

با توجه به اهمیت کیفیت آب به‌عنوان یکی از معیارهای اصلی در برنامه‌ریزی منابع آب، بررسی طبقه‌بندی کیفی آب و تعیین پارامترهای تأثیرگذار در طبقه‌بندی کیفیت آب رودخانه می‌تواند کمک شایانی به تصمیم‌گیرندگان در مدیریت کیفی آب رودخانه‌ها کند. از طرفی استفاده از روش‌های متداول و مرسوم شاخص‌های کیفی آب برای ارزیابی و پایش کیفیت آب‌های سطحی پرهزینه و زمان‌بر است از این‌رو استفاده از شیوه‌های مدل‌سازی مبتنی بر روش‌های داده‌مبنا مانند PNN، یک ابزار قدرتمند برای غلبه بر این چالش‌ها و شتاب بخشیدن به پاسخ تصمیم‌گیرندگان در مواجه شدن با مشکلات و محدودیت‌ها است. در مقاله حاضر با استفاده از مدل PNN، رویکردی جدید برای طبقه‌بندی کیفی آب پیشنهاد شده است تا بتوان با حداقل پارامترهای کیفی طبقه کیفیت آب را تعیین کرد. استفاده از این روش

کیفیت برخی از رودخانه‌های واقع در دامنه‌های جنوبی کوه سهند پرداختند. نتایج ایشان نشان داد که روش تصمیم‌گیری درختی قادر است با استفاده از چهار پارامتر هدایت الکتریکی (EC)، مجموع کاتیون‌ها، نسبت جذب سدیم (SAR) و کل مواد جامد محلول (TDS) کلاس کیفیت آب را با دقت بالایی تخمین بزند. Towler و همکاران (۲۰۰۹) با استفاده از روش KNN^۱ کیفیت آب را شبیه‌سازی کردند. نتایج این پژوهش نشان داد که روش مورد استفاده در عین سادگی و انعطاف‌پذیری، توانایی بالایی در تشخیص صحیح کیفیت بسیاری از نمونه‌های مورد بررسی داشته است. Singh و همکاران (۲۰۱۱) با استفاده از مدل‌های SVM و رگرسیون بردار پشتیبان (SVR)^۲ به مدیریت کیفیت آب‌های سطحی پرداختند. در این مطالعه طبقه‌بندی مکانی و زمانی ۱۵۰۰ نمونه به گروه‌های مشابه از نظر کیفیت آب و توسعه یک مدل SVR مناسب برای پیش‌بینی BOD^۳ با استفاده از مجموعه‌ای از متغیرها انجام گرفت. نتایج تحقیقات ایشان نشان داد که در طراحی مجدد مدل SVR، نیاز به داده‌های کمتری در حدود ۹۲/۵ درصد داده‌های قبل دارد. همچنین ایشان بیان کردند، مدل SVR یک ابزار مناسب برای پیش‌بینی BOD با استفاده از چند متغیر قابل اندازه‌گیری است. Modaresi و Araghinejad (۲۰۱۴) با استفاده از سه مدل شامل، شبکه عصبی احتمالاتی (PNN)^۴، KNN و SVM به طبقه‌بندی کیفیت آب آبخوان تهران با استفاده از اطلاعات ۱۰۰ چاه مشاهده‌ای پرداختند. در این تحقیق ابتدا با استفاده از شاخص کیفیت آب کانادایی (CCME) کیفیت آب در ۱۰۰ چاه مشاهده‌ای در آبخوان تهران بر اساس دو پارامتر کلر و نیترات در سه کلاس عالی، بسیار بد و بد طبقه‌بندی شدند.

ایشان با استفاده از دو معیار ارزیابی عملکرد نرخ خطا (ER)^۵ و مقدار خطا (EV)^۶ کارایی مدل‌های ذکر شده را بررسی کردند. نتایج نشان داد که SVM، در مراحل واسنجی و اعتبار سنجی بهترین و الگوریتم KNN با

¹ K-Nearest Neighbor

² Support Vector Regression

³ Biochemical Oxygen Demand

⁴ Probabilistic Neural Network

⁵ Error Rate

⁶ Error Value

⁷ Discriminant Analysis

⁸ Naïve Bayesian Classification

اروند رود قرار دارد. در حال حاضر حجم بسیار بالایی از زه آب کشاورزی و فاضلاب صنعتی به ویژه فاضلاب صنایع سلولزی و غذایی به این محیط‌های آبی تخلیه می‌گردد. لذا این امر موجب می‌شود که نه تنها کیفیت آب شرب بلکه حتی کیفیت آب برای مصارف کشاورزی و صنعتی نیز با مشکلات جدی مواجه گردد. در این تحقیق به منظور طبقه‌بندی کیفیت آب این رودخانه، از آمار کیفی ۱۷۲ نمونه آب (در بازه زمانی سال‌های ۱۳۸۶ تا ۱۳۹۱) در ایستگاه‌های گتوند، عرب اسد، شوشتر، زرگان، اهواز، ملاثانی، کوت امیر و دارخوین استفاده شده است. مشخصات این ایستگاه‌ها در جدول ۱ و موقعیت جغرافیایی آن‌ها در شکل ۱ ارائه شده است.

باعث کاهش قابل توجه هزینه‌های نمونه‌برداری و آزمایشگاهی شده و این امکان را برای کارشناسان و محققان فراهم می‌سازد تا در مدت زمان کوتاه و تعداد پارامتر کیفی کمتری طبقه کیفیت آب را تعیین نمایند. بدین منظور در پژوهش حاضر، در کنار مدل PNN از شاخص NSFQI در تعیین طبقه کیفیت آب رودخانه کارون استفاده شد و نتایج مورد بررسی قرار گرفت.

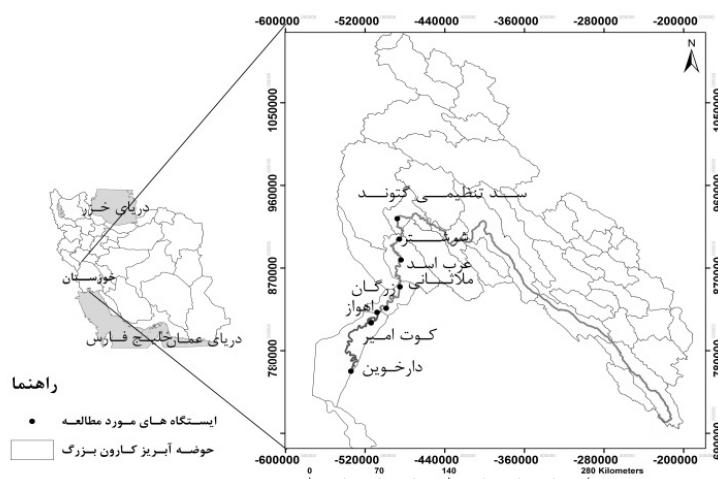
۲. مواد و روش‌ها

۱.۲. معرفی منطقه مورد مطالعه

استان خوزستان در جنوب غربی ایران با مساحتی حدود ۶۴۲۳۶ کیلومترمربع در کنار خلیج فارس و

جدول ۱. موقعیت ایستگاه‌های مورد مطالعه

نام رودخانه	ایستگاه	طول جغرافیایی	عرض جغرافیایی
کارون	گتوند	۴۸° ۴۹'	۳۲° ۱۴' ۵۶"
شطیط	عرب اسد	۴۸° ۵۲' ۵۰"	۳۱° ۵۱' ۴۵"
گرگر	شوشتر	۴۸° ۵۱' ۰۰"	۳۲° ۰۱' ۵۵"
کارون	زرگان	۴۸° ۴۵' ۴۱"	۳۱° ۲۳' ۲۶"
کارون	اهواز	۴۸° ۴۱' ۴۱"	۳۱° ۲۰' ۱۶"
کارون	ملاثانی	۴۸° ۵۲' ۴۰"	۳۱° ۳۵' ۰۱"
کارون	کوت امیر	۴۸° ۳۶' ۱۶"	۳۱° ۱۳' ۵۴"
کارون	دارخوین	۴۸° ۲۵' ۳۸"	۳۰° ۴۵' ۰۱"



شکل ۱) موقعیت رودخانه کارون و ایستگاه‌های مورد مطالعه

پارامترها به معیارهای ۱۰۰-۰ تبدیل می‌شوند. در این روش برای محاسبه شاخص نهایی هر یک از زیر شاخص‌های به دست آمده از منحنی‌های مربوطه در فاکتور وزنی خود ضرب شده و از حاصل جمع آن‌ها طبق رابطه ۱ شاخص نهایی به دست می‌آید. جدول ۲ فاکتورهای وزنی به کار رفته برای هر پارامتر را در شاخص NSFQI نشان می‌دهد. در نهایت شاخص به دست آمده بر اساس جدول ۳ به صورت کیفی طبقه‌بندی می‌شود (Abbasi et al., 2012).

$$NSFWQI = \sum W_i I_i \quad (1)$$

که در این رابطه I_i = زیر شاخص I_i و W_i = ضریب وزنی زیر شاخص I_i است.

جدول ۳. رده بندی کیفی آب بر اساس مقادیر شاخص NSFQI

شاخص محاسبه شده وضعیت کیفیت آب	
عالی	۹۱-۱۰۰
خوب	۷۱-۹۰
متوسط	۵۱-۷۰
بد	۲۶-۵۰
بسیار بد	۰-۲۵

طبقه بندی یاد گرفته شده استفاده می‌شود (Modaresi & Araghinejad, 2014). شبکه‌های عصبی احتمالی الگوها را بر اساس استراتژی بیزی و تخمین زنده‌های غیرپارامتریک توابع چگالی احتمال طبقه بندی می‌کنند و از جمله روش‌های هستند که مبتنی بر الگوهای آماری موجود در داده عمل می‌کنند (Masters, 1995). استراتژی بیزی به مجموعه‌ای از قواعد و استراتژی‌هایی اطلاق می‌شود که به منظور طبقه بندی

۲.۲. شاخص NSFQI

در سال ۱۹۷۰ با حمایت بهداشت ملی آمریکا، براون و همکاران یک شاخص کیفی کاهشی را بر اساس نظرسنجی از تعداد زیادی از افراد متخصص با تخصص‌های گوناگون در این زمینه ارائه کردند (Brown et al., 1970). آن‌ها در ابتدا حدود ۳۵ پارامتر را معرفی کرده و سپس بر اساس نظر افراد متخصص حدود نه پارامتر را برای ایجاد شاخص اصلی انتخاب کردند که شامل پارامترهای زیر است: اکسیژن‌خواهی بیوشیمیایی (BOD5)، اکسیژن محلول (DO)، کلیفرم مدفوعی، نیترات، خاصیت اسیدی (pH)، دما، کل مواد جامد (TS)، فسفات کل، کدورت (Landwehr et al., 1976; Brown et al., 1972). پس از اندازه‌گیری مشخصه‌های فوق، زیر شاخص هر یک از آن‌ها از روی منحنی‌های تبدیل به دست می‌آید که با استفاده از این منحنی‌ها

جدول ۲. فاکتورهای وزنی به کار رفته در شاخص NSFQI

پارامترها	فاکتور وزنی
کدورت	۰/۰۸
BOD	۰/۱۱
DO	۰/۱۷
کلیفرم مدفوعی	۰/۱۶
نیترات	۰/۱
pH	۰/۱۱
درجه حرارت	۰/۱
TS	۰/۰۷
فسفات	۰/۱

۳.۲. شبکه‌های عصبی احتمالی (PNN)

شبکه‌های عصبی مصنوعی برای اولین بار توسط اسپیچت معرفی شدند (Specht, 1990). این گونه شبکه‌ها برای حل مسائل پیچیده همچون برآورد (تقریب)، تشخیص الگو و طبقه بندی به کار می‌روند. در این روش از یادگیری نظارتی به منظور ایجاد تابع‌های توزیع در لایه الگو استفاده می‌شود. از این توابع برای تخمین احتمال بردار ویژگی ورودی به عنوان بخشی از کلاس یا

رابطه ۳ مثالی از استفاده کرنل گوسی برای هر مشاهده متغیر تصادفی به منظور تخمین تابع تراکم آن است:

$$f_i(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sigma^p} \frac{1}{n_i} \sum_{k=1}^{n_i} e^{-\frac{(X-X_{i,k})^T(X-X_{i,k})}{2\sigma^2}} \quad (3)$$

در این معادله، x بردار ورودی، k تعداد متغیرهای موجود در بردار ورودی، n_i تعداد الگوهای آموزشی موجود در کلاس i ام، $X_{i,k}$ k امین الگوی آموزشی موجود در کلاس i ام و σ پارامتر هموارسازی است. تابع چگالی احتمال فوق‌الذکر با به کارگیری روش تقریب پارزن مستقیماً از مجموعه داده‌های آموزشی برآورد می‌شود (Cacoulios, 1966; Wasserman, 1993). PNN یک الگوریتم نظارتی است، لذا پیش از استفاده برای طبقه‌بندی نیاز به آموزش دارد. در مرحله صحت‌سنجی، پارامتر هموارسازی باید با استفاده از سعی و خطا مشخص شود تا خطاهای طبقه‌بندی در بردارهای آموزشی کاهش یابد.

۴.۲. مدل‌سازی جهت تعیین طبقه‌بندی کیفیت آب

به منظور مدل‌سازی تعیین طبقه کیفیت آب توسط PNN ابتدا یک مجموعه از ۱۷۲ نمونه آب و طبقه کیفیت آن‌ها به عنوان ورودی و خروجی‌های مدل تعیین شد، به این صورت که پارامترهای کیفیت به عنوان ورودی مدل و کلاس‌های کیفیت آب حاصل از شاخص NSFQI، خروجی مدل در نظر گرفته شدند. ۲۵٪ از مجموعه داده‌های ورودی و خروجی به عنوان داده‌های مجموعه آزمایش^۴ و بقیه به عنوان مجموعه آموزش^۵ انتخاب گردید. آموزش شبکه با استفاده از داده‌های آموزش و کلاس‌های تعیین شده برای هر نمونه انجام شد. برای

الگوها از کمینه کردن ریسک مورد انتظار استفاده می‌کنند (Kim et al., 2005). یک تکنیک طبقه‌بندی الگو در حالت کلی با استفاده از یک مجموعه از ورودی‌های n بعدی، به تصمیم‌گیری در مورد تعلق یک مشاهده به یک دسته خاص می‌پردازد. قاعده تصمیم‌گیری بیزی برای طبقه‌بندی یک الگوی مشخص (x) و تعلق آن به یک طبقه خاص عبارت است از:

$$x \in c_r \text{ if } h_r l_r f_r(x) \geq h_s l_s f_s(x) \quad (2)$$

for $\forall s, s \neq r, 1 \leq r \leq q$

به قسمی که l_i ضرر متناظر با تعلق اشتباه الگوی x به کلاس i ام، h_i یک احتمال اولیه از تعلق الگوی x به کلاس i ام و $f_i(x)$ تابع چگالی احتمال کلاس i ام برای ورودی x است (Khashei et al., 2013).

ساختار شبکه‌های عصبی احتمالی در حالت کلی شامل چهار لایه است. یک لایه ورودی و سه لایه پردازش اطلاعات که شامل لایه الگو^۱، لایه کلاس‌بندی و لایه خروجی است. نورون‌های لایه ورودی^۲ برابر با تعداد فاکتورهای ورودی است. نورون‌های این لایه صرفاً وظیفه انتقال مقادیر ورودی به تمام نورون‌های لایه دوم را بر عهده داشته و هیچ‌گونه پردازشی در این لایه صورت نمی‌گیرد (Yang et al., 2008). در لایه الگو تعداد کل نورون‌ها برابر با مجموع تعداد نورون‌های استفاده شده برای نشان دادن الگوها در هر کلاس هستند. هر کلاس می‌تواند شامل تعداد زیادی الگوی آموزشی (بردارهای آموزشی) باشد که تعداد آن برابر با فاکتورهای ورودی است. تابع فعال‌سازی در لایه الگو را می‌توان از برخی توابع چگالی کرنل انتخاب کرد اما غالباً از کرنل گوسی استفاده می‌شود. در لایه جمع‌بندی^۳، تعداد نورون‌ها برابر با تعداد کلاس‌ها است. علاوه بر این، تابع فعال‌سازی در این لایه، یک وزن‌دار ساده است. در نهایت در لایه خروجی، تنها یک نورون وجود دارد که کلاس و طبقه ورودی مورد نظر را مشخص می‌کند (Chen et al., 2003; Modaresi & Araghinejad, 2014).

¹ Pattern Layer

² Input Layer

³ Summation Layer

⁴ Testing set

⁵ Training set

دقت نیز نشان دهنده تعداد پیش‌بینی‌های درست است. این معیارها به صورت زیر تعریف می‌شوند:

$$ER = \frac{a}{a+b} \times 100 \quad (۴)$$

$$EV = (class_o - class_s)^2 \quad (۵)$$

$$Acc = \frac{b}{a+b} \times 100 \quad (۶)$$

$$R^2 = 1 - \frac{6(\sum d_i^2)}{n(n^2 - 1)} \quad (۷)$$

در روابط بالا، a تعداد کلاس‌های نادرست پیش‌بینی شده و b تعداد کلاس‌های درست پیش‌بینی شده است. Class o کلاس کیفی مشاهده شده و Class s کلاس کیفی شبیه‌سازی شده، n تعداد داده‌ها و d_i^2 اختلاف بین رتبه‌های دو متغیر است.

۶.۲. تعیین پارامترهای تأثیرگذار در طبقه‌بندی

کیفی آب

به منظور تعیین پارامترهای تأثیرگذار در طبقه‌بندی آب رودخانه با استفاده از مدل PNN، ابتدا در هر مرحله، هر یک از نه پارامتر در نظر گرفته شده حذف شدند و تأثیر حذف هر پارامتر بر روند تعیین طبقه کیفی آب بررسی شد. برای بررسی دقیق‌تر مدل، ترکیب دوتایی پارامترها، سه‌تایی و... آن‌ها حذف شدند و عملکرد مدل با استفاده از معیارهای ارزیابی مورد بررسی قرار گرفت. در ادامه نتایج حاصل شده گزارش شده‌اند.

۳. نتایج

همان‌طور که پیش از این تشریح شد در تحقیق حاضر، عملکرد الگوریتم PNN در تعیین طبقه‌بندی کیفیت آب در مقایسه با شاخص NSFQI مورد ارزیابی قرار گرفت. برای این منظور ابتدا با استفاده از شاخص

جلوگیری از بیش‌برازش^۱ مدل، از صحت‌سنجی متقاطع^۲ استفاده گردید. در فرآیند صحت‌سنجی متقاطع در هر بار اجرای مدل با یک پارامتر مشخص، یک داده از مجموعه آموزش به عنوان داده صحت‌سنجی کنار گذاشته می‌شود و مدل براساس داده‌های باقی‌مانده، آموزش می‌بیند، سپس داده کنار گذاشته شده به مجموعه آموزش برگردانده می‌شود و داده دیگری از داده‌های آموزش به عنوان داده صحت‌سنجی کنار گذاشته می‌شود و بار دیگر شبکه آموزش داده می‌شود. این فرآیند تا زمانی که تمام داده‌های آموزش در قسمت صحت‌سنجی استفاده شوند ادامه می‌یابد. پس از آن بر اساس میانگین خطای مدل در تخمین داده‌های صحت‌سنجی، پارامتر مناسب شبکه تعیین می‌شود. در این مرحله، پارامتر پهنای^۳ که تعیین‌کننده پارامتر σ در الگوریتم PNN است مشخص می‌شود (Rumpf et al., 2010; Xue et al., 2005). مقدار بهینه این پارامتر نقشی مهمی در عملکرد الگوریتم PNN و جلوگیری از بیش‌برازش دارد. در این تحقیق مقدار بهینه این پارامتر با سعی و خطا و بر اساس صحت‌سنجی متقاطع برابر ۳۵۰ به دست آمد.

۵.۲. معیارهای ارزیابی مدل

برای ارزیابی کارایی الگوریتم PNN، از معیارهای ارزیابی نرخ خطا (ER)، مقدار خطا (EV)، دقت^۴ (Acc) و ضریب همبستگی اسپیرمن^۵ در دو مرحله آموزش و آزمایش استفاده شد. نرخ خطا نشان‌دهنده تعداد داده‌هایی است که به اشتباه طبقه‌بندی شده‌اند و مقدار خطا نشان‌دهنده مقدار خطای دسته‌بندی در داده‌هایی است که به اشتباه دسته‌بندی شده‌اند. روشن است که مقدار صفر این دو معیار بیانگر عدم وجود خطا در طبقه‌بندی و مقادیر بزرگتر از صفر وجود خطا در طبقه‌بندی را نشان می‌دهد.

^۱ Over Training

^۲ Cross Validation

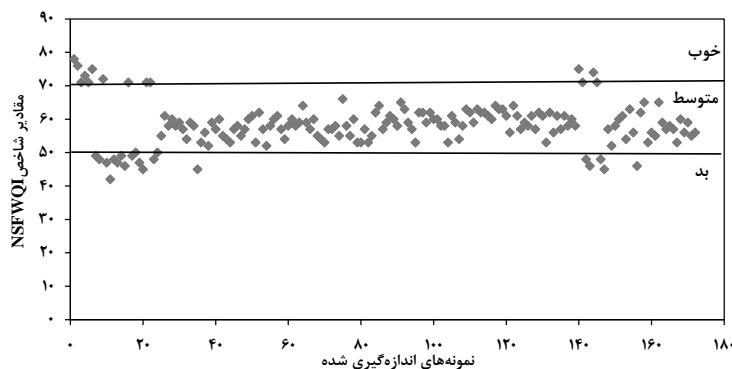
^۳ Spread

^۴ Accuracy

^۵ Spearman correlation coefficient

طبقه‌بندی نمونه‌های مورد بررسی را بر اساس شاخص NSFQWI نشان می‌دهد.

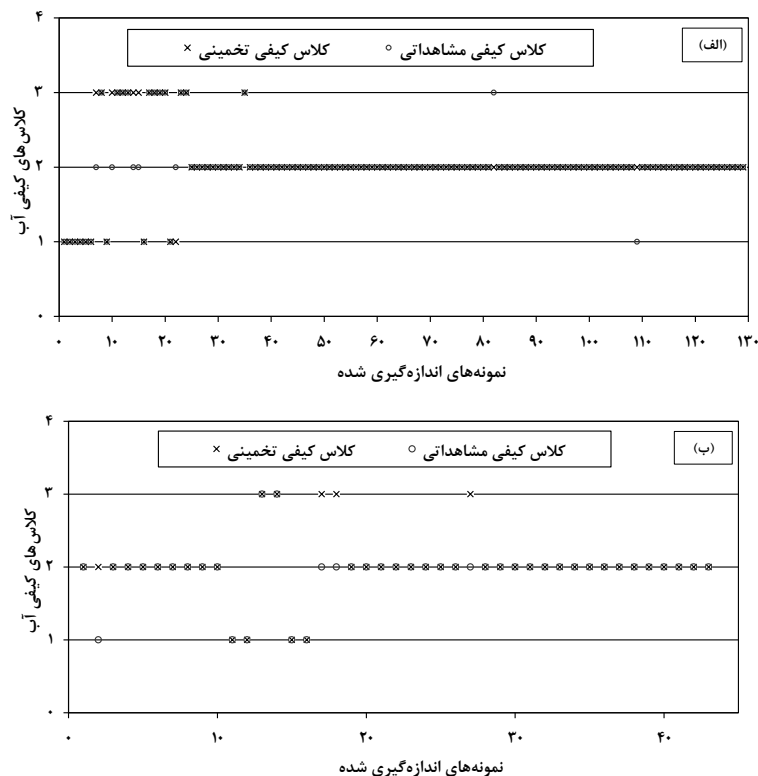
NSFWQI، طبقه‌بندی کیفی نمونه‌های مورد بررسی انجام گرفت. نتایج به دست آمده نشان داد که نمونه‌ها در سه کلاس کیفی خوب، بد و متوسط قرار گرفتند. شکل ۲



شکل ۲) طبقه‌بندی داده‌های مشاهداتی بر اساس شاخص NSFQWI

یک از نه پارامتر در نظر گرفته شده حذف شدند و تأثیر حذف هر پارامتر بر روند پیش‌بینی طبقه‌بندی آب بررسی شد. جدول ۵ نتایج حاصل از حذف تک پارامترها را نشان می‌دهد. همان‌طور که مشخص است با حذف پارامترهای کدورت و کل مواد جامد، معیار ارزیابی نرخ خطا در مرحله آموزش از ۵/۴۳ درصد به ۷/۷۵ درصد افزایش یافته و دقت طبقه‌بندی از ۹۴/۵۷ درصد به ۹۲/۲۵ درصد کاهش یافته است. حذف پارامتر کلیفرم مدفوعی نیز باعث ایجاد خطای زیادی در طبقه‌بندی شد، به طوری که درصد نرخ خطا در مراحل آموزش و آزمایش به ترتیب به ۱۷/۸۳ درصد و ۲۳/۲۶ درصد افزایش یافته و دقت طبقه‌بندی به ۸۲/۱۷ درصد و ۷۶/۷۴ درصد به ترتیب در مراحل آموزش و آزمایش کاهش یافته است. همچنین حذف این پارامتر باعث شده که ۲۳ و ۱۰ نمونه از ۱۲۹ و ۴۳ نمونه در نظر گرفته شده برای آموزش آزمایش با یک اختلاف کلاس به اشتباه طبقه‌بندی شوند. برای بررسی دقیق‌تر الگوریتم، ترکیب دوتایی پارامترها و بعد از آن ترکیب سه‌تایی و... آن‌ها حذف شدند.

شکل ۳ نتایج حاصل از طبقه‌بندی کیفی آب را با استفاده از مدل PNN در مرحله آموزش و آزمایش نشان می‌دهد. در این الگوریتم طبقه یک، دو و سه به ترتیب بیانگر طبقه کیفی خوب، متوسط و بد هستند. از ۱۲۹ و ۴۳ داده‌ای که به ترتیب برای مرحله آموزش و آزمایش در نظر گرفته شدند ۷ و ۴ نمونه با یک اختلاف کلاس به اشتباه طبقه‌بندی شده‌اند، لذا مقدار خطا برابر ۷ و ۴ در مراحل آموزش و آزمایش محاسبه شد. ضریب همبستگی محاسبه شده بین کلاس کیفی مشاهداتی و کلاس کیفی محاسبه شده در مراحل آموزش و آزمایش به ترتیب برابر با ۰/۸۵ و ۰/۷۷ حاصل شد که نشان‌دهنده همبستگی بالا بین کلاس مشاهداتی و کلاس محاسبه شده توسط PNN است. نتایج حاصل از معیارهای ارزیابی بیانگر آن است که PNN در مرحله آموزش و آزمایش به ترتیب از دقت ۹۴/۵۷ درصد و ۹۰/۷۰ درصد برخوردار است که بیانگر عملکرد خوب الگوریتم در تعیین کلاس‌های کیفی آب است. به طور خلاصه نتایج حاصل از ارزیابی مدل در مراحل آموزش و آزمایش در جدول ۴ آورده شده است. به منظور تعیین پارامترهای تأثیرگذار در طبقه‌بندی آب رودخانه با استفاده از PNN، ابتدا در هر مرحله، هر



شکل ۳. نتایج حاصل از طبقه بندی کیفی آب با استفاده از PNN (الف) مرحله آموزش (ب) مرحله آزمایش

جدول ۴. نتایج حاصل از معیارهای ارزیابی در مرحله آموزش و آزمایش PNN

ضریب همبستگی		نرخ خطا (%)		مقدار خطا		دقت (%)	
آموزش	آزمایش	آموزش	آزمایش	آموزش	آزمایش	آموزش	آزمایش
۰/۸۵۲	۰/۷۷۳	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰

جدول ۵. نتایج حاصل از حذف یک پارامتر بر روند تعیین طبقه بندی کیفیت آب با استفاده از الگوریتم PNN

پارامتر حذف شده	نرخ خطا (%)		مقدار خطا		دقت (%)	
	آموزش	آزمایش	آموزش	آزمایش	آموزش	آزمایش
pH	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
کدورت	۷/۷۵	۹/۳۰	۱۰	۴	۹۲/۲۵	۹۰/۷۰
فسفات	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
دما	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
BOD	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
DO	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
نیترات	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰
کلیفرم مدفوعی	۱۷/۸۳	۲۳/۴۶	۲۳	۱۰	۸۲/۱۷	۷۶/۷۴
کل مواد جامد	۷/۷۵	۹/۳۰	۱۰	۴	۹۲/۲۵	۹۰/۷۰

پارامترهای pH، فسفات، دما، BOD، DO و نیترات عملکرد مدل به ترتیب در مراحل آموزش و آزمایش با نرخ خطای ۵/۴۳ درصد و ۹/۳۰ درصد، مقدار خطای ۷ و ۱۰ و دقت ۹۴/۵۷ درصد و ۹۰/۷۰ درصد در تعیین کلاس کیفی آب تغییر نکرده است.

جدول ۶، نتایج عملکرد PNN را از حذف ترکیبی تعدادی از پارامترها نشان می‌دهد. مقایسه معیارهای ارزیابی به دست آمده از حذف پارامترهای کیفی مختلف بیانگر آن است که تنها پارامترهای کلیفرم مدفوعی، کل مواد جامد و کدورت در تعیین طبقه‌بندی کیفیت آب مؤثرند. همان‌طور که در جدول ۶ مشخص است با حذف

جدول ۶. نتایج حاصل از حذف چند پارامتر بر روند تعیین طبقه‌بندی کیفیت آب با استفاده از الگوریتم PNN

پارامتر حذف شده		نرخ خطا (%)		مقدار خطا		دقت (%)	
آموزش	آزمایش	آموزش	آزمایش	آموزش	آزمایش	آموزش	آزمایش
کدورت و pH	۷/۷۵	۹/۳۰	۱۰	۴	۹۲/۲۵	۹۰/۷۰	
کدورت و کلیفرم مدفوعی	۲۰/۹۲	۱۹/۳۸	۲۵	۹	۸۰/۶۲	۷۹/۰۷	
فسفات و DO	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰	
دما و BOD	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰	
کلیفرم مدفوعی و کل مواد جامد	۱۷/۸۲	۲۳/۲۶	۲۳	۱۰	۸۲/۱۷	۷۶/۷۴	
DO و نیترات	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰	
کدورت، فسفات و pH	۷/۷۵	۹/۳۰	۱۰	۴	۹۲/۲۵	۹۰/۷۰	
کدورت، BOD و کل مواد جامد	۱۰/۰۸	۹/۳۰	۱۳	۴	۸۹/۹۲	۹۰/۷۰	
کلیفرم مدفوعی، دما و DO	۱۷/۸۲	۲۳/۲۶	۲۳	۱۰	۸۲/۱۷	۷۶/۷۴	
نیترات، DO، BOD، فسفات، دما، pH	۵/۴۳	۹/۳۰	۷	۴	۹۴/۵۷	۹۰/۷۰	

هر ماه اندازه‌گیری کنند و سایر پارامترها هر چند ماه یکبار اندازه‌گیری شوند.

به طور کلی نتایج نشان داد که استفاده از PNN، روش مناسبی برای جایگزینی شاخص‌های کیفیت آب جهت طبقه‌بندی کیفی آب و ارزیابی پارامترهای کیفی آب است. همان‌طور که پیش از این گفته شد، PNN یک مدل نظارت شده است که از مهم‌ترین ویژگی‌های این مدل‌ها این است که اندازه شبکه مورد نیاز و نرخ خطای طبقه‌بندی به طور مستقیم در روند ساختار شبکه گنجانیده شده است، در نتیجه الگوریتم پیشنهادی اغلب دارای یک ساختار شبکه کوچک و دقت طبقه‌بندی قابل قبول است (Mao et al., 2000).

نتایج حاصل از حذف پارامترهای کیفی آب نشان می‌دهد حذف پارامترهای کدورت و کل مواد جامد ۲/۴۶ درصد دقت طبقه‌بندی را کاهش داده است. همچنین حذف پارامتر کلیفرم مدفوعی باعث شده دقت به ترتیب در مراحل آموزش و آزمایش ۱۳/۱۱ درصد و ۱۸/۸۵ درصد کاهش یابد، بنابراین پارامتر کلیفرم مدفوعی مؤثرترین پارامتر در مشخص کردن کلاس کیفی آب تعیین شد، زیرا با حذف این پارامتر خطای زیادی در طبقه‌بندی کیفی آب ایجاد شد. با توجه به اینکه هزینه‌های سالانه پایش در تعیین پارامترهای کیفی آب امری مهم و درخور توجه است PNN قادر است با استفاده از سه پارامتر کیفی کدورت، کل مواد جامد و کلیفرم مدفوعی طبقه‌بندی کیفی آب را تعیین کند، لذا توصیه می‌شود که برای مشخص کردن طبقه کیفی آب، پارامترهای کلیفرم مدفوعی، کل مواد جامد و کدورت را

۴. بحث و نتیجه گیری

در این تحقیق، به بررسی و ارزیابی کیفیت آب‌های سطحی مربوط به ۱۷۲ نمونه آب رودخانه کارون در هشت ایستگاه با استفاده از شاخص NSFQI و مدل PNN پرداخته شد. تعیین طبقه بندی کیفی آب با استفاده از شاخص های کیفی آب هزینه بر و زمان بر است؛ از این رو در این مطالعه، استفاده از مدل PNN، به عنوان یک جایگزین برای شاخص های سنتی کیفیت آب، به کار گرفته شد. در این مدل می توان با استفاده از حداقل پارامترهای کیفی، طبقه کیفی آب را مشخص کرد. نتایج پایش کیفی نمونه ها با استفاده از شاخص NSFQI، بیانگر آن بود که نمونه ها در سه طبقه خوب، بد و متوسط قرار دارند. برای آموزش مدل PNN ترکیب مختلف پارامترهای کیفی به عنوان ورودی و نتیجه طبقه کیفی حاصل از شاخص NSFQI به عنوان خروجی استفاده شد. در مقایسه با شاخص NSFQI، الگوریتم PNN تنها با استفاده از سه پارامتر کیفی کدورت، کلیفرم مدفوعی و کل مواد جامد با دقت ۹۴/۵۷٪ و ۹۰/۷۰٪ در مراحل آموزش و آزمایش عملکرد قابل قبولی را در تعیین طبقه بندی کیفی آب نشان داد. مقایسه نتایج این تحقیق با سایر تحقیقات نشان دهنده سازگاری نتایج تحقیق حاضر با تحقیق های پیشین است. برای مثال Xue و همکاران (۲۰۰۵) در پژوهش خود برای طبقه بندی ترکیبات فعال در گیاهان دارویی نشان دادند، که به دلیل اینکه PNN تنها یک پارامتر قابل تنظیم دارد، ساختاری ساده تر و سریع تر در مقایسه با شبکه های عصبی با انتشار به عقب^۱ (BP) دارند، بنابراین یک روش خوب برای طبقه بندی به شمار می رود. Balabin و همکاران (۲۰۱۰) در تحقیق خود که به طبقه بندی سوخت

پرداختند، به این نتیجه رسیدند که از میان نه روش طبقه بندی مختلف در نظر گرفته شده، مدل PNN بهترین نتایج را ارائه کرده است. Modaresi و Araghinejad (۲۰۱۴) در تحقیق خود نشان دادند که الگوریتم PNN یک روش طبقه بندی مناسب با آموزش سریع برای ارزیابی کیفیت آب است. Khaki و همکاران (۲۰۱۵) در شبیه سازی پارامترهای TDS و EC با استفاده از سایر مقادیر پارامترهای کیفی آب توسط شبکه های عصبی مصنوعی^۲ (ANN) و سیستم فازی-عصبی^۳ نشان دادند که رویکردهای هوش مصنوعی و سیستم های فازی-عصبی قادر به تفسیر رفتار پارامترهای کیفی آب هستند. Sakizadeh و Mirzaei (۲۰۱۶) با استفاده از دو مدل SVM و KNN در طبقه بندی آب زیرزمینی آبخوان خوزستان به این نتیجه رسیدند که پارامترهای نترات و کلسیم تأثیرگذارترین پارامترها در طبقه بندی هستند. نتایج تحقیق حاضر نیز مانند مطالعات گذشته توانایی روش های داده مبنا را در طبقه بندی و پیش بینی کیفیت آب نشان می دهد. بر این اساس که نتایج در تمام تحقیقات مناسب بوده است، می توان بیان کرد که طبقه بندی کیفی آب با استفاده از این روش ها، مسئله پیچیده ای محسوب نمی شود و این مدل ها به راحتی قابلیت حل مسئله را دارند. بنابراین استفاده از الگوریتم PNN باعث کاهش هزینه های نمونه برداری و زمان محاسبات برای تعیین کلاس کیفیت آب شده است و جایگزینی مناسب برای شاخص کیفی NSFQI است. در مطالعات آتی پیشنهاد می شود که از الگوریتم های دیگر برای تعیین طبقه بندی کیفی آب استفاده شود و پارامترهای کیفی تأثیرگذار در طبقه بندی کیفی آب با استفاده از آنها تعیین شود.

References

Abbasi, T., Abbasi, S.A., 2012. Water quality indices. Elsevier, 375 p.

¹ Back-propagation

² Artificial Neural Network

³ Neuro-fuzzy System

- Asgari, H., Kerachian, R., 2007. Application of probability estimation of the support vector machines in water quality classification of rivers (Case study: Karoon-dez river). 23-24 Jan., Isfahan University of Technology, Esfahan, Iran. (In Persian)
- Balabin, R.M., Safieva, R.Z., Lomakina, E.I., 2010. Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques. *Analytica Chimica Acta* 671(1), 27-35.
- Brown, R.M., McClelland, N.I., Deininger, R.A., Tozer, R.G., 1970. A water quality index- do we dare. *Water and Sewage Works*, 339-343.
- Brown, R.M., McClelland, N.I., Deininger, R.A., O'Connor, M.F., 1972. A water quality index—crashing the psychological barrier. In: *Indicators of environmental quality*. Springer, US, pp. 173-182.
- Cacoullos, T., 1966. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics* 18(1), 179-189.
- Chen, A.S., Leung, M.T., Daouk, H., 2003. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research* 30(6), 901-923.
- Horton, R.K., 1965. An index number system for rating water quality. *Journal of Water Pollution Control Federation* 37(3), 300-306.
- Khaki, M., Yusoff, I., Islami, N., 2015. Application of the Artificial Neural Network and Neuro-fuzzy System for Assessment of Groundwater Quality. *CLEAN—Soil, Air, Water* 43(4), 551-560.
- Khashei, M., Bijari, M., Mokhatab Rafeiei, F., 2013. Exchange rate forecasting using Hybrid models of multi-layer perceptrons (MLPs) and probabilistic neural network (PNNs). *Journal of Numerical Modeling in Engineering* 32(1), 97-113. (In Persian)
- Kim, D.K., Lee, J.J., Lee, J.H., Chang, S.K., 2005. Application of probabilistic neural networks for prediction of concrete strength. *Journal of Materials in Civil Engineering* 17(3), 353-362.
- Landwehr, J.M., Deininger, R.A., 1976. A comparison of several water quality indexes. *Journal (Water Pollution Control Federation)*, 954-958.
- Mao, K.Z., Tan, K.C., Ser, W., 2000. Probabilistic neural-network structure determination for pattern classification. *Neural Networks, IEEE Transactions on* 11(4), 1009-1016.
- Masters, T., 1995. *Advanced algorithms for neural networks: a C++ sourcebook*. John Wiley & Sons.
- Modaresi, F., Araghinejad, S., 2014. A Comparative assessment of support vector machines, probabilistic neural networks, and K-Nearest neighbor algorithms for water quality classification. *Water resources management* 28(12), 4095-4111.
- Rahman, M., Das, R., Hassan, N., Roy, K., Haque, F., Akber, A., 2014. Environmental study on water quality of Mayur River with reference to suitability for irrigation. *International Journal of Environmental Sciences* 5(2), 291-308.
- Rasinezami, S., Nazariha, M., Baghvand, A., Moridi, A., 2012. Karkheh river water quality using multivariate statistical analysis and qualitative data variations. *Journal of health system research* 8(7), 1280-1292. (In Persian)
- Rumpf, T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., Plümer, L., 2010. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture* 74(1), 91-99.
- Sakizadeh, M., 2015. Assessment the performance of classification methods in water quality studies, a case study in Karaj River. *Environmental monitoring and assessment* 187(9), 1-12.
- Sakizadeh, M., Mirzaei, R., 2016. A comparative study of the performance of K-nearest neighbors and support vector machines for classification of groundwater. *Journal of Mining and Environment* 7(2), 149-164.
- Sattari, M., Abbasgoli Naebzad, M., Mirabbasi Najafabadi, R., 2014. Surface water quality prediction using decision tree method. *Journal of irrigation and water engineering* 4(15), 76-88. (In Persian)
- Singh, K.P., Basant, N., Gupta, S., 2011. Support vector machines in water quality management. *Analytica chimica acta* 703(2), 152-162.
- Specht, D.F., 1990. Probabilistic neural networks. *Neural networks* 3(1), 109-118.

- Towler, E., Rajagopalan, B., Seidel, C., Summers, R.S., 2009. Simulating ensembles of source water quality using a K-nearest neighbor resampling approach. *Environmental science & technology* 43(5), 1407-1411.
- Walker, D., Jakovljević, D., Savić, D., Radovanović, M., 2015. Multi-criterion water quality analysis of the Danube River in Serbia: A visualisation approach. *Water research* 79, 158-172.
- Wasserman, P.D., 1993. *Advanced methods in neural computing*. Van Nostrand Reinhold, New York.
- Xue, C.X., Zhang, X.Y., Liu, M.C., Hu, Z.D., Fan, B.T., 2005. Study of probabilistic neural networks to classify the active compounds in medicinal plants. *Journal of pharmaceutical and biomedical analysis* 38(3), 497-507.
- Yang, Y., Ge, S.S., Lee, T.H., 2008. Hand gesture recognition and tracking based on distributed locally linear embedding. *Image and Vision Computing* 26(12), 1607-1620.