



## Evaluating the performance of advanced machine learning models in predicting the Urmia Lake water level

R. Hajian<sup>1</sup>, M.R. Jalali<sup>2\*</sup>, and R. Mastouri<sup>3</sup>

### Abstract

Lakes play an important role in the hydrological cycle and predicting the water level in them can provide vital information for the future management of lakes and their ecosystems. To this aim, in this research two models were developed and run for 1-, 2- and 3-month water level forecasts for the Lake Urmia in north-west of Iran; the tree post-pruning using the REPT method and the combined REPT model with the ROF-REPT. The water level time series data from 2001 to 2020 were divided into two categories for model building (from 2001 to 2014) and for model validation (from 2015 to 2020). Different input scenarios were developed and evaluated to find the most effective input scenario of climate variables. Finally, the developed models were evaluated through visual and quantitative criteria. The results showed that the combined ROF-REPT model has a higher performance than the single REPT model for all forecasts of 1-, 2- and 3-months. Nash-Sutcliffe Efficiency was obtained between 0.45 and 0.87 for single models and between 0.53 and 0.95 for combined models. Also, it was shown that the developed models are able to predict the water level up to 3 months ahead.

**Keywords:** Water Level, Lake Urmia, Decision Tree Models, Hybrid Model, Iran.

Received: September 2, 2022

Accepted: November 21, 2022

## ارزیابی مدل‌های پیشرفته یادگیری ماشین در پیش‌بینی تراز سطح آب دریاچه ارومیه

رسول حاجیان<sup>۱</sup>، محمدرضا جلالی<sup>۲\*</sup> و رضا مستوری<sup>۳</sup>

### چکیده

دریاچه‌ها نقش مهمی در چرخه هیدرولوژیکی دارند و پیش‌بینی سطح آب آنها می‌تواند اطلاعات حیاتی برای مدیریت آینده دریاچه‌ها و اکوسیستم آنها فراهم کند. در پژوهش حاضر ۲ مدل شامل پس هرس کردن درخت به روش کاهش خطای هرس و مدل ترکیبی REPT با مدل جنگل چرخان (ROF-REPT) توسعه و ساخته شد، و برای پیش‌بینی ۱، ۲ و ۳ ماه آبی سطح آب دریاچه ارومیه در شمال غرب ایران مورد استفاده قرار گرفت. داده‌های سری زمانی سطح آب از سال ۲۰۰۱ تا ۲۰۲۰ به دو دسته، به ترتیب برای ساخت مدل (از سال ۲۰۰۱ تا ۲۰۱۴) و اعتبارسنجی (از ۲۰۱۵ تا ۲۰۲۰) تقسیم شد. سناریوهای ورودی و رودی مختلف برای یافتن مؤثرترین سناریو ورودی از متغیرهای اقلیمی ساخته شد و مورد ارزیابی قرار گرفت. در نهایت مدل‌های توسعه‌یافته از طریق معیارهای بصری و کمی ارزیابی شدند. نتایج نشان داد که مدل ترکیبی ROF-REPT دارای عملکرد بالاتری نسبت به مدل منفرد REPT برای تمامی ۱، ۲ و ۳ ماه آینده است. ضریب نش (Nash-Sutcliffe Efficiency) برای مدل‌های منفرد بین ۰/۴۵ تا ۰/۸۷ و برای مدل‌های ترکیبی بین ۰/۵۳ تا ۰/۹۵ حاصل شد. همچنین، نشان داده شد که مدل‌های توسعه‌یافته قادر به پیش‌بینی سطح آب تا ۳ ماه آینده هستند.

**کلمات کلیدی:** تراز سطح آب، دریاچه ارومیه، مدل‌های درخت تصمیم، مدل ترکیبی، ایران.

تاریخ دریافت مقاله: ۱۴۰۱/۶/۱۱

تاریخ پذیرش مقاله: ۱۴۰۱/۸/۳۰

1- PhD Candidate, Department of Civil Engineering, Arak branch, Islamic Azad University, Arak, Iran. Email: 30vil.ir@gmail.com

2- Assistant Professor, Department of Civil Engineering, Arak branch, Islamic Azad University, Arak, Iran. Email: mr.jalali@iau.ac.ir

3- Assistant Professor, Department of Civil Engineering, Arak branch, Islamic Azad University, Arak, Iran. Email: reza.mastoori@gmail.com

\*- Corresponding author

Dor: [20.1001.1.17352347.1402.19.2.3.0](https://doi.org/10.17352/347.1402.19.2.3.0)

۱- دانشجوی دکتری مدیریت منابع آب، دانشکده فنی و مهندسی واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

۲- استادیار گروه عمران، دانشکده فنی و مهندسی واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

۳- استادیار گروه عمران، دانشکده فنی و مهندسی واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

\*- نویسنده مسئول  
بحث و مناظره (Discussion) در مورد این مقاله تا پایان تابستان ۱۴۰۲ امکانپذیر است.



## ۱- مقدمه

معتدل منطقه تبدیل به هوای گرمسیری با بادهای نمکی خواهد شد و زیست محیط منطقه تغییر خواهد کرد. علاوه بر نمک، بسیاری از آلودگی‌های شامل فلزات سمی سنگین مورد استفاده در صنعت و مواد سمی مورد استفاده در کشاورزی به آب‌های سطحی و زیر سطحی مرتبط با دریاچه نفوذ کرده‌اند و در صورت خشک شدن دریاچه بسیاری از مواد سمی هوایی شده و خطرات بیماری‌های تنفسی برای زیست‌بوم و مردم منطقه به وجود خواهد آورد. به طور کلی، پیش‌بینی تراز سطح آب دریاچه برای ماه‌های آتی در مدیریت جامع منابع آب دریاچه بسیار مؤثر و مهم است (Aghakochek et al., 2015).

از آنجا که نوسانات سطح آب دریاچه‌ها به طور غیر خطی به بسیاری از تغییرات طبیعی و انسانی پاسخ می‌دهد (به عنوان مثال، باد، مدیریت مخزن و غیره)، مدل‌های غیرخطی برای پیش‌بینی دقیق سطح آب دریاچه‌ها معمولاً به کار گرفته می‌شوند (Zhu et al., 2020). بر این اساس، در سال‌های اخیر، انواع مختلفی از مدل‌های مبتنی بر الگوریتم عددی، آماری و یادگیری ماشین برای مدل‌سازی نوسانات سطح آب دریاچه‌ها استفاده شده است (Huang et al., 2010; Khatibi et al., 2014). با این حال، روش‌های عددی (به عنوان مثال، مدل process-oriented (Fry et al., 2020))، مدل برنامه مسیریابی برای محاسبه موازی دبی (RAPID) (Lin et al., 2015) و مدل‌های آماری، محدودیت‌هایی را برای پیش‌بینی تراز سطح آب دریاچه دارند: (۱) مدل‌های عددی علاوه بر اینکه وقت‌گیر هستند، (۲) به مجموعه داده‌های گسترده و کاملی از جمله باسیمتری دریاچه، متغیرهای هواشناسی، ورودی‌ها، خروجی‌ها و غیره نیاز دارند (Zhu et al., 2020) و (۳) کالیبراسیون آنها نیز بسیار دشوار است. از طرف دیگر مدل‌های آماری رگرسیون خطی که فرضیات بیش از حد محدودکننده را بر داده‌ها تحمیل می‌کنند، دارای یک ساختار خطی ساده هستند. مدل‌های یادگیری ماشین (ML) که در سراسر جهان برای پیش‌بینی تراز سطح آب دریاچه‌ها مورد استفاده قرار می‌گیرند: (۱) دارای یک ساختار کلی بوده، (۲) غیر پارامتری هستند و (۳) به دلیل ساختار انعطاف‌پذیر و غیرخطی، می‌توانند روابط غیرخطی بین داده‌های ورودی و خروجی را تخمین بزنند.

Ahmadi Nasab et al. (2013) به پیش‌بینی تراز آب دریاچه ارومیه با استفاده از روش برنامه‌ریزی ژنتیک پرداختند. با استفاده از داده‌های روزانه تراز سطح آب دریاچه ارومیه از سال ۱۳۴۵ تا سال ۱۳۹۲، اقدام به پیش‌بینی تراز سطح ایستایی دریاچه با استفاده از روش برنامه‌ریزی ژنتیک تا سال ۱۳۹۵ کردند. آنها بیان داشتند که مدل موجود دارای

دریاچه ارومیه، بزرگ‌ترین دریاچه داخلی ایران، بزرگ‌ترین دریاچه آب شور در خاورمیانه، و ششمین دریاچه بزرگ آب شور دنیا است. این دریاچه با داشتن بیش از یک صد جزیره کوچک صخره‌ای محل توقف پرندگان مهاجر از جمله فلامینگو، پلیکان، کفچه‌نوک، اکراس، لک‌لک، اردک پیسه‌نوک خنجری، چوب‌پا و مرغ نوروزی است. به خاطر شوری بیش از حد دریاچه هیچ نوع ماهی در این دریاچه زندگی نمی‌کند. با این حال، دریاچه ارومیه یکی از زیست‌گاه‌های مهم سخت‌پوست آرمیا شناخته می‌شود. این سخت‌پوست یکی از منابع اصلی تغذیه پرندگان مهاجر از جمله فلامینگو به‌شمار می‌آید. دریاچه ارومیه در سال ۱۳۵۲ در فهرست پارک‌های ملی ایران به ثبت رسید. نوسانات سطح آب دریاچه‌ها، عامل اصلی کنترل عملکرد اکوسیستم‌های آنها است. بررسی نوسانات سطح آب دریاچه‌هایی مانند ارومیه که بسیار شور هستند، علاوه بر اینکه در تخصیص آب، پیش‌بینی دقیق سلامت اکوسیستم‌ها، توسعه پایدار و مدیریت منابع آب درون حوضه‌ای و نیز توسعه اقتصادی ضروری و مهم است، در تعیین خشکی آنها که تبعات جبران‌ناپذیری از قبیل منبع اصلی برای ریزگردها و به‌ویژه پراکنده شدن نمک در استان‌های همجوار دریاچه به دنبال خواهد داشت نیز اهمیت دارد. تراز آب دریاچه ارومیه طی سال‌های اخیر با کاهش جدی مواجه شده است که برای آن دلایل مختلفی را مانند تغییرات اقلیمی و عوامل انسانی بیان کرده‌اند. طی یک مقاله مروری از حدود ۵۴۴ مقاله چاپ شده طی ۴۰ سال اخیر بیان کردند که کشاورزی، سدسازی و سوء مدیریت نقش بیشتری از اقلیم و تغییرات اقلیمی در خشک شدن دریاچه ارومیه داشته است. Parsinejad et al. (2022) بیان کردند که با ترکیبی جدید از مشاهدات ماهواره‌ای چندطیفی با وضوح بالا، نشان می‌دهیم که مساحت این دریاچه ایران در دهه‌های گذشته حدود ۸۸ درصد کاهش یافته است که بسیار بیشتر از آنچه قبلاً گزارش شده بود می‌باشد (۲۵ تا ۵۰ درصد). آنها بیان کردند که خط ساحلی دریاچه به شدت در حال عقب‌نشینی است و هیچ نشانه‌ای از بهبود وجود ندارد، که بخشی از آن به دلیل خشکسالی‌های طولانی مدت است. همچنین عنوان کردند که ما از داده‌های آب و هوای تنظیم‌شده با سنجنده ماهواره‌ای حوضه دریاچه از داده‌های شاخص بارش استاندارد استفاده می‌کنیم تا نشان دهیم که عقب‌نشینی در حال انجام خط ساحلی صرفاً مصنوعی خشکسالی‌های طولانی مدت نیست. تغییرات شدید در سلامت دریاچه‌ها در درجه اول پیامدهای برنامه‌های توسعه منابع آب منطقه‌ای تهاجمی، ناشی از فعالیت‌های کشاورزی فشرده، تغییرات انسانی در سیستم و رقابت بالادستی بر سر آب است. کارشناسان محیط زیست و منابع آب پیش‌بینی می‌کنند در صورت خشک شدن این دریاچه هوای

دریاچه ارومیه توسط مدل ترکیبی شبکه عصبی مصنوعی و مدل سری زمانی میانگین متحرک پرداختند. Vaheddoost and Aksoyi. (2021) از متغیرهای بارش، دبی، تراز آب زیرزمینی و تبخیر به عنوان ورودی به مدل سازی استفاده کردند. آنها بیان کردند که ترازهای قبلی آب دریاچه مهمترین متغیر ورودی برای مدل سازی بوده و همچنین مدل ساخته شده آنها دارای دقت قابل قبولی است. اگرچه مدل های منفرد یادگیری ماشین با موفقیت در مدل سازی هیدرولوژیکی به کار گرفته شده اند، اما ایجاد مدل های ترکیبی با (الف) افزودن روش های پیش پردازش مانند تبدیل موجک به مخلوط کردن یا گرفتن تغییرات چند مقیاسه در داده های سری زمانی (Quilty and Adamowski, 2016)، (ب) بهبود ساختار مدل های یادگیری ماشین از طریق اجرای الگوریتم های تکاملی (به عنوان مثال، الگوریتم ژنتیک، بهینه سازی ازدحام ذرات PSO)، یا (پ) نمونه برداری مجدد از مجموعه داده ها، برای کاهش خطا و واریانس در خروجی مدل با استفاده از روش های ترکیبی (به عنوان مثال، Staking, boosting, و رگرسیون افزایشی (AR<sup>8</sup>) (Khosravi et al., 2022)، عملکرد پیش بینی را نسبت به مدل های منفرد افزایش می دهد. در پژوهش حاضر، از روش های گروه BA برای توسعه مدل های ترکیبی استفاده خواهد شد. از آنجا که روش BA از نمونه گیری مجدد با فرایند جایگزینی برای تولید مجموعه های مختلف آموزش تصادفی برای به دست آوردن یک یادگیرنده منحصر به فرد استفاده می کند (Tiwari and Chatterjee, 2011). برای مدل سازی هیدرولوژیکی مناسب است، به ویژه هنگامی که مجموعه داده های آموزش کوچک باشد (Barzegar et al., 2020). همچنین، در پژوهش های قبلی به تأثیر متغیرهای ورودی به مدل سازی و تأثیر آنها و همچنین ساخت بهترین ترکیب ورودی برای مدل سازی به خوبی پرداخته نشده است. به علاوه، در بیشتر مطالعات فقط به پیش بینی تراز آب دریاچه از طریق انطباق داده های حال حاضر استفاده شده است، و برای پیش بینی برای افق های آینده به درستی انجام نشده است. بنابراین، هدف از پژوهش حاضر: (۱) تعیین پارامترهای اقلیمی مؤثر بر مدل سازی تراز سطح آب دریاچه ارومیه، (۲) تعیین مؤثرترین ترکیب ورودی از متغیرهای ورودی به مدل، (۳) افزایش دقت مدل سازی برای پیش بینی تراز سطح آب دریاچه ارومیه به عنوان یک سیستم دارای آشوبناکی (Hassanzadeh et al., 2013) از طریق ترکیب مدل ها و (۴) ساخت مدلی برای پیش بینی تراز آب سطح دریاچه ارومیه برای ۱ تا ۳ ماه آینده با کمترین مقدار داده است.

کارایی مناسبی برای پیش بینی تراز آب دریاچه ارومیه است (Ahmadi Nasab et al., 2013). (Komasi et al. 2015) به مقایسه سه روش پیش بینی سری زمانی آریمای، شبکه عصبی مصنوعی و شبکه عصبی- موجکی جهت ارائه بهترین روش پیش بینی تراز سطح آب دریاچه ارومیه در مقیاس ماهانه پرداختند. بر مبنای نتایج حاصل شده، عملکرد شبکه عصبی- موجکی در مقایسه با دو روش سری زمانی آریمای و شبکه عصبی مصنوعی دارای دقت بالاتری بوده است. Bahmanesh et al. (2015) به پیش بینی سطح تراز دریاچه ارومیه در بازه زمانی ۱۳۴۵ تا ۱۳۸۷ با استفاده از مدل سری های زمانی خطی پرداختند. بیان کردند که مدل حاضر دارای کارایی بالایی در پیش بینی تراز آب دریاچه ارومیه است. (Karimzadegan et al. 2017) به مقایسه کاربرد روش های شبکه بیزین و حداقل مربعات ماشین بردار پشتیبان در پیش بینی تراز سطح آب دریاچه ارومیه پرداختند. آنها بیان کردند که مقایسه و بررسی نتایج دو مدل با بررسی ضرایب  $R^2$ ، RMSE،  $MBE^4$  و  $NSE^5$  حاکی از برتری عملکرد مدل حداقل مربعات ماشین بردار پشتیبان در مقایسه با شبکه بیزین است. آنها به پیش بینی تراز آب دریاچه ارومیه برای ۱ و ۷ روز آتی توسط مدل ماشین یادگیری حداکثری، برنامه ریزی بیان ژن و شبکه عصبی مصنوعی پرداختند. آنها از متغیرهای تراز سطح آب دریاچه شامل تراز میانگین، حداکثر تراز و حداقل تراز به عنوان متغیرهای ورودی به مدل استفاده کردند. Shiri et al. (2017) بیان کردند که مدل ماشین یادگیری حداکثری دارای قدرت پیش بینی بالاتری است. آنها به پیش بینی تراز سطح آب دریاچه تیتیکاکا در آمریکای جنوبی پرداختند. آنها بیان کردند که به منظور استفاده بهینه از دریاچه ها، پیش بینی دقیق سطح آب دریاچه موضوع اصلی مدیریت منابع آب است. آنها از ترکیب مدل رگرسیون بردار پشتیبان (SVR<sup>6</sup>) و الگوریتم گرگ خاکستری (GWO<sup>7</sup>) برای پیش بینی نوسانات سطح آب دریاچه استفاده کردند. همچنین برای یافتن بهترین متغیرهای ورودی برای پیش بینی تراز سطح آب دریاچه توسط مدل های SVR و SVR-GWO از سه نوع روش پیش پردازش داده ها یعنی تحلیل مؤلفه های اصلی، جنگل تصادفی و الگوریتم تسکین استفاده کردند. قبل از شبیه سازی تراز سطح آب دریاچه در مرحله زمانی ماهانه با استفاده از مدل ترکیبی، یک رویکرد تکاملی مبتنی بر تاخیرهای ماهانه مختلف برای تعیین بهترین ترکیب متغیرهای ورودی انجام شد. نتایج نشان داد که بر اساس روش جنگل تصادفی، بهترین سناریو ورودی  $X_{t-1}$ ،  $X_{t-2}$ ،  $X_{t-3}$ ،  $X_{t-4}$  برای مدل SVR-GWO بود. همچنین عملکرد مدل SVR-GWO نشان داد که می تواند تراز سطح آب دریاچه را با دقت قابل قبولی شبیه سازی کند (Mohammadi et al., 2020). آنها به پیش بینی تراز آب

## ۲- منطقه مورد مطالعه

در این پژوهش، دریاچه ارومیه که در شمال غربی ایران بین استان‌های آذربایجان شرقی و آذربایجان غربی در ایران و در غرب بخش جنوب‌غربی دریای خزر واقع شده است (شکل ۱) به عنوان منطقه و دریاچه مورد مطالعه انتخاب گردید. مساحت دریاچه ارومیه ۵۲۰۰ کیلومتر مربع، طول آن ۱۴۰ کیلومتر، عرض آن ۵۵ کیلومتر و حداکثر عمق آن ۱۶ متر است. در اواخر سال ۲۰۱۷، به دلیل خشکسالی، بهره‌برداری بیش از حد از آب‌های زیرزمینی، ساخت سد و سایر سازه‌های مصنوعی مانند ساخت جاده‌های غیرمهندسی بر فراز دریاچه، دریاچه به ۱۰ درصد از اندازه قبلی خود کوچک شد.

## ۳- روش شناسی

### ۳-۱- داده‌های مورد استفاده

در پژوهش حاضر تأکید بر تأثیر متغیرهای اقلیمی به عنوان پارامترهای مؤثر بر نوسانات تراز سطح آب دریاچه است. در این تحقیق داده‌های ماهانه شامل بارش (R)، دبی‌های ورودی به دریاچه (Q)، تبخیر (E) و تراز سطح آب (WL) از ۲۰۰۱/۱/۱ تا ۲۰۲۰/۱۲/۳۱ از شرکت آب منطقه‌ای آذربایجان غربی جمع‌آوری و به دست آمد. داده‌های تبخیر در ایستگاه گامانخانه آبشور و داده‌های بارش در ایستگاه سینوپتیک ارومیه به عنوان ایستگاه نزدیک به دریاچه ارومیه جمع‌آوری شدند.

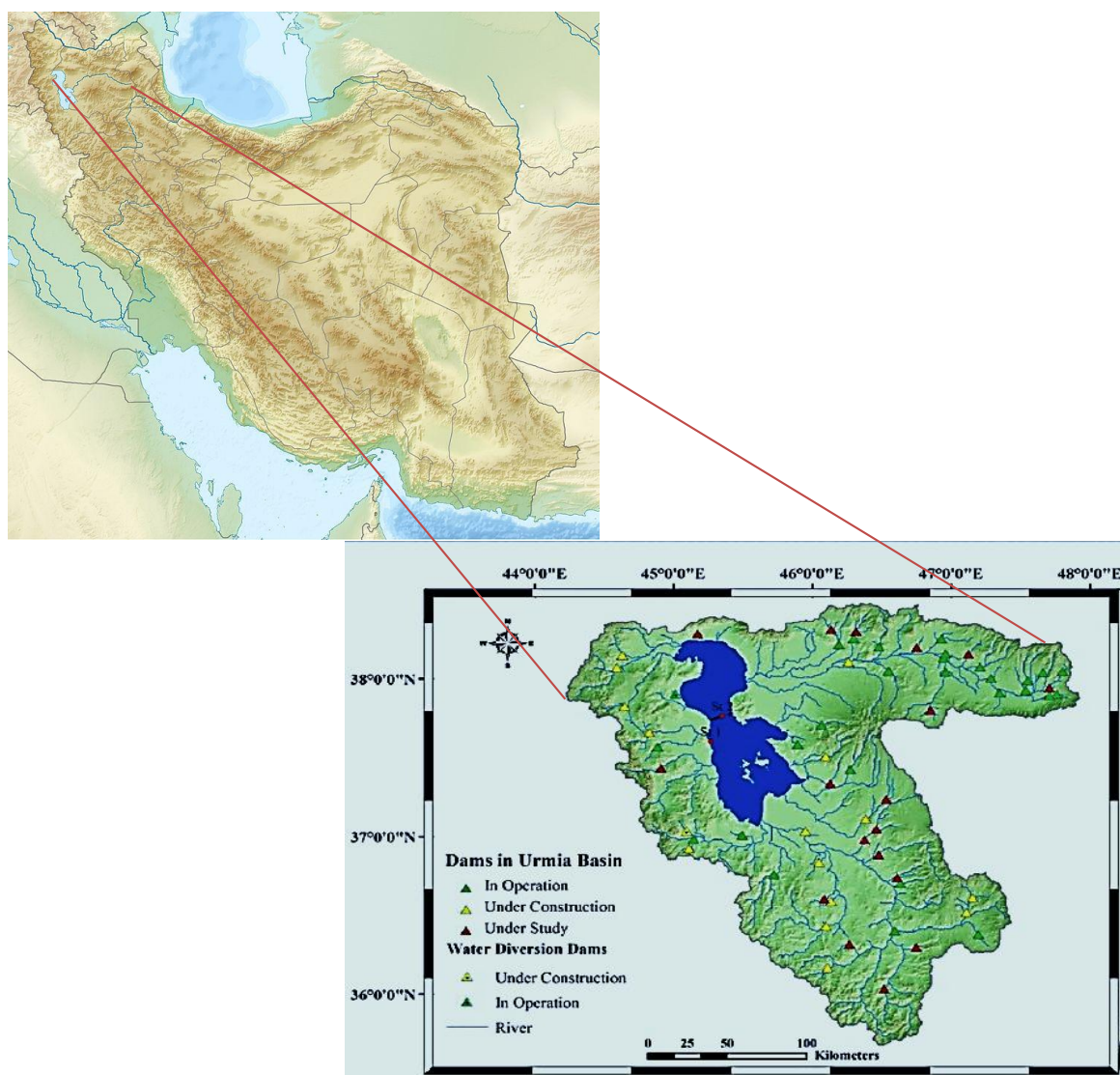


Fig. 1- The location of Lake Urmia in Iran and the location of the dams in the watershed of the lake (Mohabzadeh et al., 2018)

شکل ۱- موقعیت دریاچه ارومیه در ایران به همراه موقعیت سدهای موجود در حوزه آبخیز دریاچه (Mohabzadeh et al., 2018)

(E) قادر به پیش‌بینی تراز سطح آب دریاچه با دقت بالا است یا خیر. در مدل‌سازی از طریق یادگیری ماشین، یافتن مدلی که قادر به پیش‌بینی متغیر هدف با دقت بالا و تنها با استفاده از تعداد متغیرهای کمتر و در دسترس‌تر باشد، می‌تواند به عنوان مؤثرترین مدل در نظر گرفته شود.

### ۳-۳- مقادیر بهینه پارامترهای مدل

علاوه بر کیفیت داده‌ها، پیچیدگی داده‌ها، قدرت پیش‌بینی مدل و متغیرهای ورودی مرتبط، یافتن مقدار بهینه برای پارامتر هر مدل تأثیر قابل‌توجهی بر نتیجه مدل‌سازی دارد. هیچ مقدار بهینه‌ای برای پارامتر هر مدل وجود ندارد که در همه موقعیت‌ها به خوبی کار کند و بنابراین، یافتن مؤثرترین مقادیر، یک مرحله حیاتی در مدل‌سازی با استفاده از مدل‌های یادگیری ماشین است. در پژوهش حاضر از رویکرد آزمون و خطا برای یافتن پارامتر بهینه مدل استفاده شده است. در مرحله اول مقادیر پیش‌فرض در هر مدل در نظر گرفته شد و هر مدل با استفاده از مقادیر پیش‌فرض اجرا گردید. در مرحله بعد، مقادیر بالاتر و پایین‌تر از مقادیر پیش‌فرض به صورت تصادفی وارد شده و مدل اجرا شد. در نهایت، نتایج با استفاده از معیارهای ریشه میانگین مربعات خطا (RMSE) برای یافتن مقدار بهینه مقایسه و ارزیابی شد.

### ۳-۴- پیشینه تئوری مدل‌ها

#### – مدل درخت پس هرس کردن درخت به روش Reduced Error Pruning

این روش بدین صورت است که ابتدا به درخت اجازه داده می‌شود تا به اندازه کافی رشد کند. سپس گره‌هایی که باعث افزایش دقت دسته‌بندی نمی‌شوند، هرس می‌گردند. در این روش داده‌ها به دو مجموعه اعتبارسنجی و آموزش تقسیم می‌شوند و در ادامه درخت با داده‌های آموزش مطابق روش قبل یاد گرفته می‌شود.

از آنجایی که دریاچه ارومیه از رودخانه‌های متعدد تغذیه می‌شود، داده‌های ماهانه جریان رودخانه‌های اصلی تغذیه‌کننده دریاچه از جمله رودخانه‌های سیمینه‌رود، زرینه‌رود، زلاچای، باراندوزچای، نازلوچای و شهرچای جمع‌آوری و ادغام شدند و به عنوان دبی کلی برای دریاچه در نظر گرفته شد. در نهایت داده‌ها به دو قسمت تقسیم شدند. حدود ۷۰ درصد داده‌ها برای آموزش و ساخت مدل و ۳۰ درصد باقی‌مانده نیز برای اعتبارسنجی و آزمون مدل مورد استفاده قرار گرفت. شایان ذکر است که هیچگونه "قانون کلی" برای تقسیم‌بندی مجموعه آموزش و آزمون وجود ندارد. با این حال، طبق مطالعات قبلی (Barzegar et al., 2016)، زیرمجموعه آزمون باید از داده‌هایی استفاده کند که هرگز در مرحله آموزش استفاده نشده‌اند و این داده‌ها باید تقریباً از ۴۰٪-۱۰٪ اندازه مجموعه آموزش باشند. رویکرد تقسیم‌بندی مجموعه داده به ۷۰:۳۰ برای آموزش و آزمون مدل‌های یادگیری ماشین بسیار مورد استفاده قرار گرفته است (Pérez-Chacón et al., 2020) و بنابراین در این پژوهش نیز از این درصد تقسیم‌بندی استفاده شد.

### ۳-۲- متغیرهای ورودی

در این مطالعه سه سناریوی ترکیبی ورودی مختلف برای یافتن مؤثرترین سناریو ورودی برای پیش‌بینی تراز سطح آب دریاچه ارومیه ایجاد و ارزیابی شدند. ابتدا سناریوی ورودی از متغیرهای هواشناسی-هیدرولوژی (E, Q, R) ساخته و مورد آزمایش قرار گرفت. سناریوی ورودی دوم ترکیبی از تراز سطح آب دریاچه برای دوره‌های قبلی است و سناریوی سوم از ادغام دو سناریوی فوق‌الذکر ساخته شده است (جدول ۱).

برای سناریوی ورودی هواشناسی-هیدرولوژی، فرض بر آن است که مشخص شود آیا متغیر با بالاترین ضریب همبستگی (به عنوان مثال

Table 1- Different scenarios for making different input combinations

جدول ۱- سناریوهای مختلف برای ساخت ترکیب‌های ورودی مختلف

output	Input type	input
$WL_{(t+1)}, WL_{(t+2)}, WL_{(t+3)}$	Meteorological variables	E
$WL_{(t+1)}, WL_{(t+2)}, WL_{(t+3)}$	Meteorological-hydrological variables	E, Q
$WL_{(t+1)}, WL_{(t+2)}, WL_{(t+3)}$	Meteorological-hydrological variables	E, Q, R
$WL_{(t+1)}, WL_{(t+2)}, WL_{(t+3)}$	Only the water level of the lake	$WL_{(t)}, WL_{(t-1)}, WL_{(t-2)}, WL_{(t-3)}$
$WL_{(t+1)}, WL_{(t+2)}, WL_{(t+3)}$	hybrid	E, Q, R, $WL_{(t)}, WL_{(t-1)}, WL_{(t-2)}, WL_{(t-3)}$

سپس، برای یک گره داخلی (غیر برگ n) زیرشاخه n حذف می‌گردد و این زیر شاخه با یک برگ جایگزین می‌شود که به این برگ برچسب دسته‌بندی اکثر مثال‌های قرار گرفته تحت این شاخه نسبت داده می‌شود. در ادامه، عملکرد درخت بر روی مثال‌های اعتبارسنجی بررسی می‌شود و اگر درخت هرس شده عملکرد بهتر و یا مساوی با درخت فعلی داشت، از درخت هرس شده استفاده می‌شود. هرس کردن آن قدر ادامه می‌یابد تا هرس بیشتر، سودی نداشته باشد. شکل ۲ عملکرد این روش به صورت نموداری دو بعدی شامل اندازه گراف و دقت درخت را نشان می‌دهد. همان‌طور که از این شکل پیدا است، هرس کردن درخت تصمیم با استفاده از این روش توانسته است تا دقت درخت تصمیم بر روی کل داده‌های آموزش و تست را افزایش دهد.

### – الگوریتم جنگل چرخشی (RF<sup>9</sup>)

این الگوریتم توسط Rodriguez et al. (2006) ارائه شد. اگر X داده‌های آموزشی با یک ماتریس  $F \in \mathbb{N}^* \times n$  ویژگی‌های مجموعه داده و  $\omega$  برچسب کلاس Y از یک گروه کلاس شامل  $\{\omega_1, \dots, \omega_2\}$  باشد، فرض می‌گردد که یک گروه از مجموعه داده به صورت تصادفی به K زیرگروه و L درخت تصمیم در این الگوریتم به نام  $\{D_1, \dots, D_L\}$  تفکیک می‌گردد. از طرفی L و M دو پارامتری هستند که باید تعیین شوند. مدل ترکیبی در این الگوریتم توسط اعمال یک

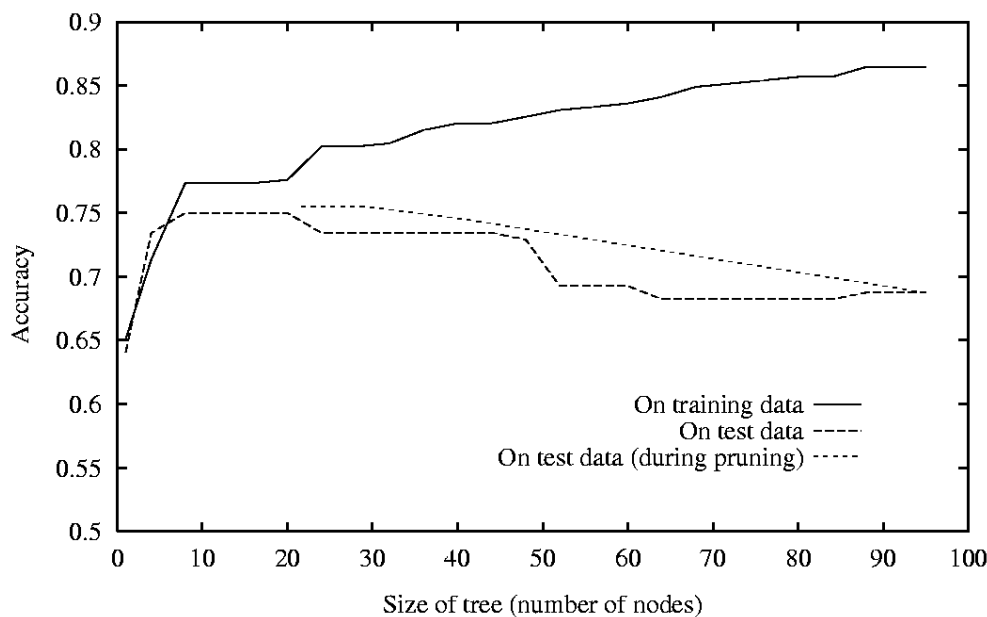
الگوریتم یادگیری پایه  $\cdot$  به انواع داده‌های آموزشی ساخته می‌شود. جهت ساختن مجموعه‌ای از داده‌های آموزشی با هدف طبقه‌بندی  $D_i$ ، نیاز به انجام مراحل زیر است:

مرحله اول: پارامتر F به طور تصادفی با هر زیر مجموعه‌ای از تعداد ویژگی‌هایش (عناصر اش)  $M=n/K$ ، به K مؤلفه تقسیم می‌شود.

مرحله دوم: اگر پارامتر  $F_{ij}$ ، z مین زیرمجموعه از مجموعه داده برای آموزش توسط الگوریتم طبقه‌بند  $D_i$  باشد و پارامتر  $X_{ij}$  مشخص کننده ویژگی داده X در پارامتر  $F_{ij}$  باشد، در این حالت، برای پارامتر  $X'_{ij}$  یک زیر مجموعه تصادفی ناتهی طراحی می‌شود و سپس یک مجموعه آموزش جدید به نام  $X'_{ij}$  که شامل مجموعه‌ای از ۷۵ درصد از این فرم از مجموعه داده‌های آموزشی براساس الگوریتم بوت استرپ<sup>۱۱</sup> است تشکیل می‌شود. در نهایت، برای تولید ضرایب ماتریس  $C_{ij}$  یک تبدیل خطی بر روی پارامتر  $X'_{ij}$  اعمال می‌شود. اندازه هر ماتریس از  $X'_{ij}$  با ضرایب  $a^{(1)}_{ij}, \dots, a^{(M_j)}_{ij}$  برابر با  $M \times 1$  است.

مرحله سوم: برای ایجاد یک ماتریس کوچک چرخشی، یعنی  $R_{ij}$ ، محاسبه ضرایب ماتریس  $C_{ij}$  به صورتی که در زیر آمده، لازم است:

$$(۱) \begin{bmatrix} a^{(1)}_{i1}, \dots, a^{(M1)}_{i1} & \{0\} \dots & \{0\} \\ \{0\} & a^{(2)}_{i2}, \dots, a^{(M2)}_{i2} \dots & \{0\} \\ \dots & \dots & \dots \\ \{0\} & \dots & a^{(K)}_{iK}, \dots, a^{(MK)}_{iK} \end{bmatrix}$$



**Fig. 2- Tree pruning by Reduced Error Pruning method**  
**شکل ۲- پس هرس کردن درخت به روش Reduced Error Pruning**

**Table 2- Optimal values for each model parameter**  
**جدول ۲- مقادیر بهینه برای پارامتر هر مدل**

Parameter	Optimal values	
	REPT	ROF
Batch-size	100	100
Number of decimal places	2	2
Maximum depth of tree	1	*
Number of Iteration	*	10
Minimum variance proportion	0.002	*
Number of folds	3	*
Initial count	0	*

### ۳-۵- ارزیابی مدل‌ها

در تحقیق مذکور از چندین معیار معروف در زمینه ارزیابی مدل‌ها شامل ضریب تبیین ( $R^2$ )، خطای جذر میانگین مربعات (RMSE)، میانگین خطای مطلق (MAE) و ضریب نش-ساتکلیف (NSE) به صورت زیر استفاده خواهد شد:

$$R^2 = \left( \frac{\sum_{i=1}^n (X_o - \bar{X}_o)(X_e - \bar{X}_e)}{\sqrt{\sum_{i=1}^n (X_o - \bar{X}_o)^2 \sum_{i=1}^n (X_e - \bar{X}_e)^2}} \right)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_e - X_o)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_e - X_o| \quad (5)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (X_e - X_o)^2}{\sum_{i=1}^n (X_e - \bar{X}_e)^2} \quad (6)$$

که در آن  $X_o$  و  $X_e$  به ترتیب مقادیر مشاهداتی و برآوردی می‌باشند. همچنین  $\bar{X}_o$  و  $\bar{X}_e$  به ترتیب میانگین مقادیر مشاهداتی و برآوردی می‌باشند.  $N$  نیز تعداد داده‌های مذکور می‌باشد. به طور کلی هرچه مقدار ضریب تبیین و ضریب نش-ساتکلیف بیشتر و ریشه میانگین مربعات خطا میانگین خطای مطلق کمتر باشد، کارایی مدل بیشتر خواهد بود.

### ۴- نتایج و بحث

#### ۴-۱- اهمیت متغیرهای ورودی

بر اساس مقدار ضریب همبستگی پیرسون ( $r$ ) که در جدول ۳ نشان داده شده است، داده‌های تراز قبلی سطح آب دریاچه تأثیر بیشتری (بیش از ۰/۹۵) بر عملکرد مدل‌سازی نسبت به داده‌های هواشناسی-

در این موقع، با توجه به مجموعه ویژگی‌های اصلی، ستون‌های پارامتر  $R_i$  دوباره مرتب می‌شوند. ماتریس چرخشی جدید مرتب شده به نام  $R_i^a$  در نظر گرفته می‌شود. سپس،  $(xR_i^a)$  به عنوان داده آموزشی تبدیل شده برای طبقه‌بند  $D_i$  مشخص خواهد شد. در نهایت، تمام الگوریتم‌های طبقه‌بندی با روش موازی آموزش خواهند یافت (Rodriguez et al., 2006). مراحل طبقه‌بندی برای یک مجموعه داده آزمایشی  $X$ ، زمانی که  $d_{ij}(xR_i^a)$  احتمال حاصل از الگوریتم طبقه‌بند  $D_i$  باشد، و نیز با این فرض که  $X$  متعلق به کلاس  $j$  ( $\omega$ ) است، ارزیابی می‌شود. سپس، برای محاسبه اطمینان یک کلاس، از روش میانگین ترکیبی به صورت معادله زیر استفاده می‌شود:

$$\mu_j^{(X)} = \frac{1}{L} \sum_{i=1}^L d_{ij}(xR_i^a), j = 1, \dots, c. \quad (2)$$

در نتیجه، ماتریس چرخشی تحلیل مؤلفه اصلی (PCA) برای هر  $K$  از زیرمجموعه متغیرهای ورودی (Airdi and Grana., 2014) جهت حفظ تغییرپذیری اطلاعات در داده‌ها، محاسبه می‌گردد. ویژگی‌های (ساختار) جدید حاصله برای طبقه‌بند مبتنا<sup>۲</sup>، با استفاده از چرخش محور  $K$  شکل می‌گیرد. الگوریتم جنگل چرخشی (RF)، سعی بر افزایش تنوع و صحت کل مجموعه داده آموزشی دارد. در الگوریتم سنتی جنگل چرخشی، روش درخت تصمیم برای چرخش زیر مجموعه‌ای از سری داده‌های آموزشی، به دلیل حساسیت آنها به چرخش ویژگی‌های (عوامل یا عناصر) محور، انتخاب می‌شده است. بنابراین، نام جنگل به دلیل نوع فرآیند آن (فرم درختی) اطلاق شده است.

شایان ذکر است، در بحث الگوریتم‌های یادگیری ماشین به عنوان بخشی از علم داده‌کاوی، تنها یک الگوریتم یادگیری که توانایی اجرای کافی فرآیند یادگیری تمامی داده‌های آموزشی را داشته باشد، وجود ندارد. به همین دلیل انتخاب الگوریتم براساس روش آزمون و خطا صورت می‌گیرد (Amasiali, 2008). در این صورت، داده آموزشی آن الگوریتم به شدت به مدلی که بر اساس آن ساخته می‌شود بستگی دارد. این وضعیت مسئله انتخاب بهترین مدل را فراهم می‌آورد. مقادیر بهینه برای پارامترهای هر مدل در جدول ۲ نشان داده شده است.

**Table 3- Correlation coefficient between input and output variables**

جدول ۳- ضریب همبستگی بین متغیرهای ورودی و خروجی

	Evaporation	Rainfall	Daby	WL <sub>(t-3)</sub>	WL <sub>(t-2)</sub>	WL <sub>(t-1)</sub>	WL <sub>(t)</sub>
WL <sub>(t+1)</sub>	-0.29	0.06	0.19	0.95	0.96	0.98	0.992
WL <sub>(t+2)</sub>	-0.26	0.09	0.20	0.94	0.95	0.96	0.98
WL <sub>(t+3)</sub>	-0.24	0.09	0.17	0.94	0.94	0.95	0.96

متغیرهای تراز سطح آب ماه‌های قبلی، عملکرد مدل‌سازی را کاهش می‌دهد.

هیدرولوژی (r کمتر از ۰/۳) دارند. همچنین با افزایش ماه‌های قبل (از ۱ به ۳) از اهمیت و اثربخشی آنها کاسته می‌شود. در میان داده‌های هواشناسی-هیدرولوژی، تبخیر (r=-۰/۲۹) نسبت به جریان (r=۰/۱۹) و بارندگی (r=۰/۰۶) مؤثرتر بوده است. این روند برای سایر ماه‌های پیش‌بینی از ۱ تا ۳ ماه پیش‌بینی صادق است، فقط برای ماه‌های طولانی‌تر، اثربخشی هر متغیر کاهش می‌یابد. البته باید در نظر داشت که ضریب همبستگی پیروان تنها میزان "همبستگی خطی" بین دو متغیر تصادفی را می‌سنجد. از این رو، برای سنجش روابط پیچیده تر و غیرخطی موجود در سیستم‌های پیچیده آبی و محیط‌زیستی قابل استفاده نیست.

**Table 4- Selecting the best-input scenario based on r-value**

جدول ۴- انتخاب بهترین سناریوی ورودی بر اساس مقدار r

	REPT		ROF-REPT	
	Education	test	Education	test
1	0.92	0.10	0.90	0.17
2	0.97	0.14	0.96	0.05
3	0.97	0.15	0.96	0.06
4	0.99	0.94	0.98	0.90
5	0.58	0.39	0.57	0.37

در برخی از مطالعات برای پیش‌بینی تراز آب دریاچه‌ها، فقط از داده‌های تراز سطح آب قبلی با تاخیر استفاده شده است (Shiri et al., 2016; Razmkhah et al., 2016). در پژوهش حاضر، اثربخشی داده‌های هواشناسی-هیدرولوژی، مورد آزمایش قرار گرفته و نشان داده شده است که چندان مؤثر نیستند. یکی از دلایل این است که تغییرات آب و هوایی، روند بارندگی و تبخیر را تغییر می‌دهد و مدل‌ها قادر به پیش‌بینی تراز سطح آب با دقت بالا نیستند. دلیل دیگری که محتمل‌تر است، این است که سدهای زیادی بر روی رودخانه‌های اصلی تغذیه‌کننده دریاچه ارومیه احداث شده‌اند و همین امر باعث می‌شود که داده‌های هیدرولوژی ورودی ارتباط مؤثری با تراز سطح آب دریاچه نداشته باشند.

#### ۳-۴- نتایج عملکرد مدل‌سازی

- برای یک ماه آینده (پیش رو)

نتیجه مدل‌سازی بر حسب نمودار پراکندگی و تغییرات زمانی برای ۱ ماه آینده در شکل ۳ و ۴ نشان داده شده است. بر اساس نتیجه، همه مدل‌های ترکیبی توسعه یافته در این تحقیق سازگاری بسیار خوبی با داده‌های مشاهده شده دارند. نتایج نشان داد، مدل ترکیبی قدرت پیش‌بینی بالاتری دارد و این ناشی از ساختار منعطف‌تر مدل ترکیبی است که ساختاری مبتنی بر استفاده از دو مدل همزمان را دارد (Onyari and Ilunga, 2013).

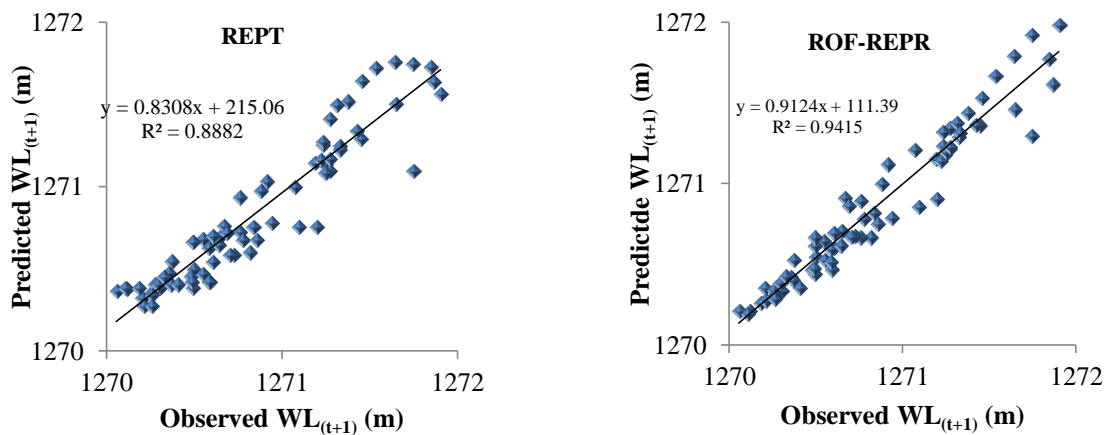
- برای دو ماه آتی (پیش رو)

نتیجه مدل‌سازی بر حسب نمودار پراکندگی و تغییرات سری زمانی برای ۲ ماه آتی در شکل ۵ و ۶ نشان داده شده است. بر اساس نتایج، مدل ترکیبی سازگاری و قدرت پیش‌بینی خوبی با داده‌های مشاهده شده، دارد. بر اساس نمودار پراکندگی ROF-REPT دارای پراکندگی کمتری در داده‌ها هست که نشان دهنده قدرت پیش‌بینی بالاتر این مدل بوده به طوری که دارای  $R^2 = 0.73$  هستند.

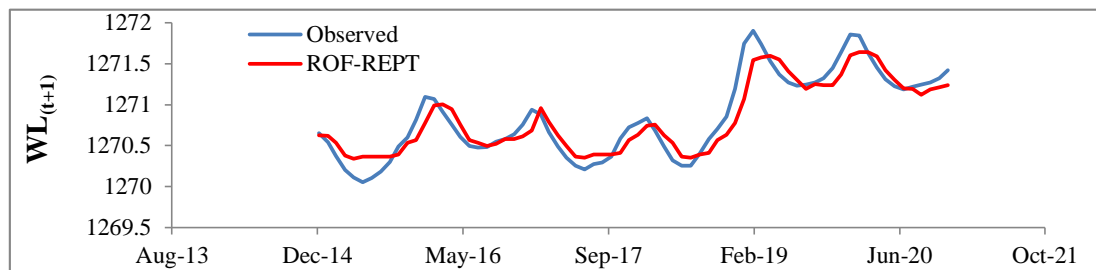
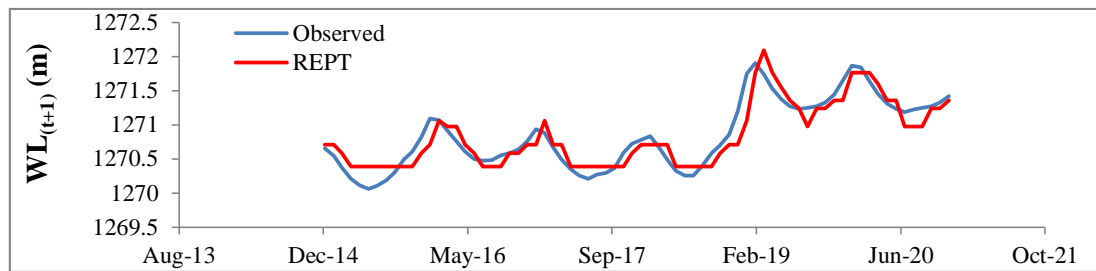
#### ۴-۲- بهترین سناریوی ورودی از متغیرها

برای همه مدل‌های توسعه یافته، سناریوی ورودی شماره ۴ که ترکیبی از تراز سطح آب برای ماه‌های قبلی است، مؤثرترین بوده و تأثیر بیشتری بر نتیجه دارد. همچنین نه متغیر هواشناسی-هیدرولوژی با بالاترین ضریب همبستگی، و نه ترکیبی از همه آنها قادر به پیش‌بینی دقیق تراز سطح آب نبودند. ورودی شماره ۵ در مقایسه با شماره ۴ نشان می‌دهد که درگیر کردن متغیر هواشناسی-هیدرولوژی با

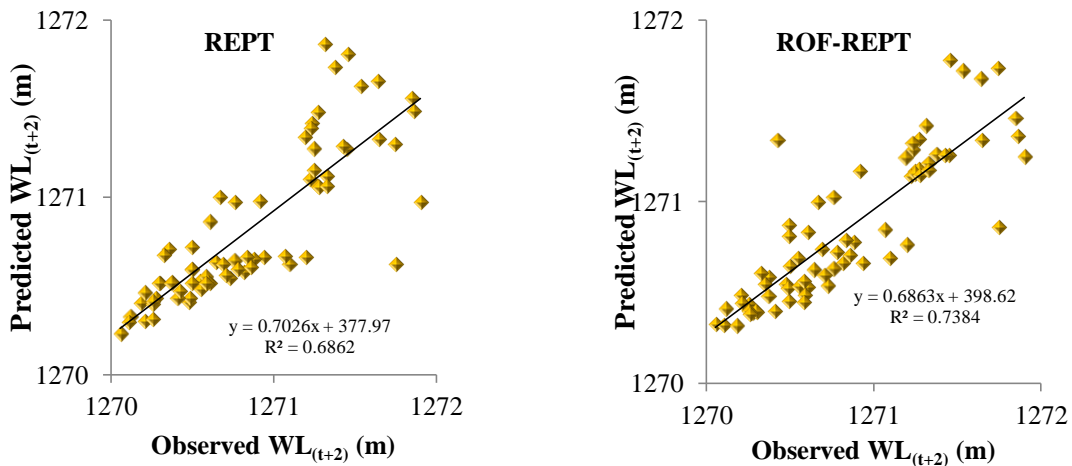




**Fig. 3- Scatter plot to evaluate modeling performance for one month ahead forecasting**  
 شکل ۳- نمودار پراکندگی برای ارزیابی عملکرد مدل سازی برای پیش بینی یک ماه آینده پیش رو



**Fig. 4- Time series for the next month in the validation period**  
 شکل ۴- نمودار سری زمانی برای یک ماه آینده پیش رو در دوره اعتبارسنجی



**Fig. 5- Scatter plot to evaluate modeling performance for two months ahead forecasting**  
 شکل ۵- نمودار پراکندگی برای ارزیابی عملکرد مدل سازی برای پیش بینی دو ماه آتی پیش رو

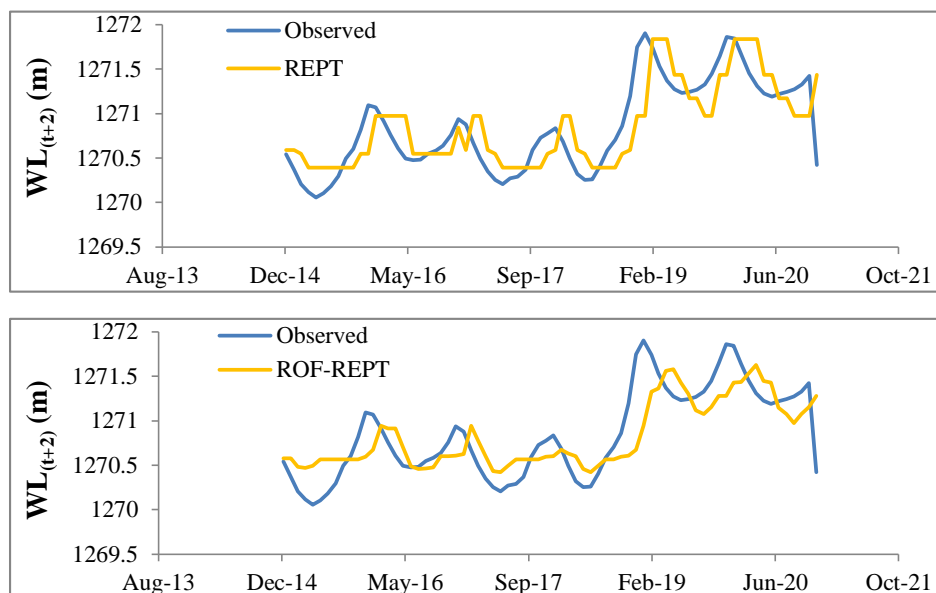


Fig. 6- Time series for the next two months in the validation period

شکل ۶- نمودار سری زمانی برای دو ماه آتی پیش‌رو در دوره اعتبارسنجی

می‌دهد. همچنین مدل منفرد REPT پراکندگی داده بیشتری دارد و در نتیجه عملکرد کمتری را نشان می‌دهد ( $R^2=0.44$ ).

نتایج داده‌های پیش‌بینی شده تراز آب دریاچه ارومیه برای یک ماه آینده با استفاده از مدل ROF-REPT با MAE, RMSE, NSE و RSR برابر با ۰/۱۱ متر، ۰/۰۸ متر، ۰/۹۵ و ۰/۳۳ دارای عملکرد بالاتری نسبت به مدل منفرد توسعه‌یافته دارد (جدول ۵).

- برای سه ماه آتی (پیش‌رو)

نتیجه مدل‌سازی بر حسب نمودار پراکندگی و تغییرات سری زمانی تراز سطح آب پیش‌بینی شده برای ۳ ماه آتی در شکل ۷ و ۸ نشان داده شده است.

بر اساس نمودار پراکندگی، ROF-REPT دارای پراکندگی داده کمتری است، که قدرت پیش‌بینی بالاترتر مدل با  $R^2=0.53$  را نشان

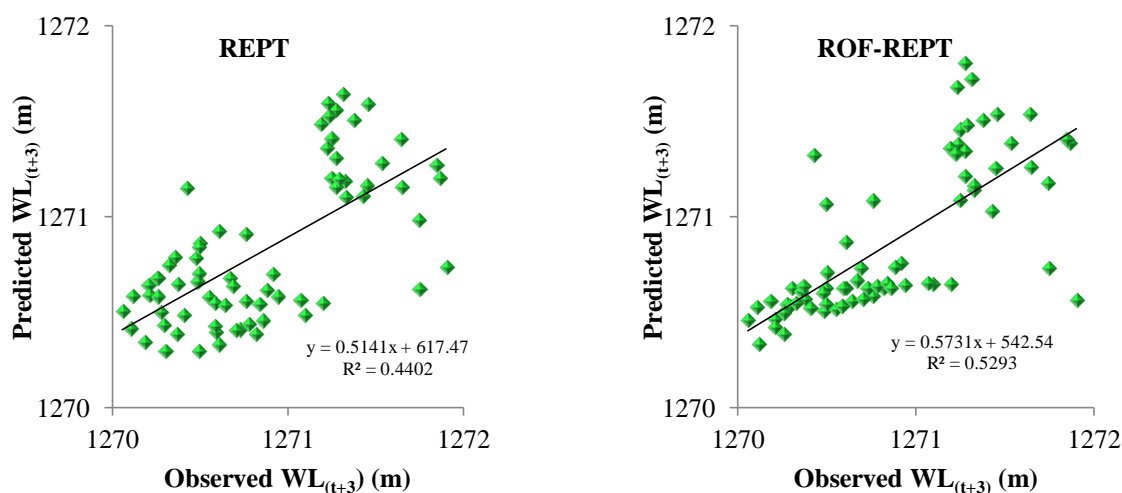


Fig. 7- Scatter plot to evaluate the modeling performance for three months ahead forecasting

شکل ۷- نمودار پراکندگی برای ارزیابی عملکرد مدل‌سازی برای پیش‌بینی سه ماه آتی پیش‌رو

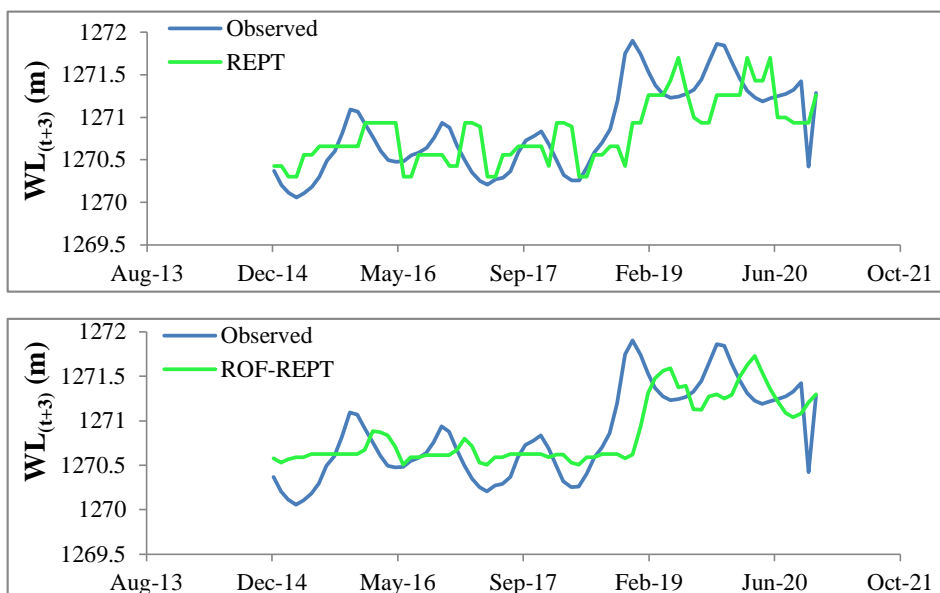


Fig. 8- Time series chart for the next three months in the validation period

شکل ۸- نمودار سری زمانی برای سه ماه آتی پیش‌رو در دوره اعتبارسنجی

Table 5- Model performance using quantitative criteria for  $WL_{(t+1)}$

جدول ۵- عملکرد مدل با استفاده از معیارهای کمی برای  $WL_{(t+1)}$

	RMSE <sup>1*</sup> (m)	MAE <sup>2*</sup> (m)	NSE <sup>3*</sup>	RSR <sup>4*</sup>
ROF_REPT	0.11	0.08	0.95	0.23
REPT	0.18	0.15	0.87	0.35

<sup>1\*</sup> Root mean square error

<sup>2\*</sup> Mean Absolute error

<sup>3\*</sup> Nash-Sutcliffe Efficiency

<sup>4\*</sup> Ratio of RMSE to the standard deviation of the observations

Table 6- Model performance using quantitative criteria for  $WL_{(t+2)}$

جدول ۶- ارزیابی عملکرد مدلها با استفاده از معیارهای کمی برای  $WL_{(t+2)}$

	RMSE (m)	MAE (m)	NSE	RSR
ROF_REPT	0.26	0.19	0.76	0.52
REPT	0.28	0.21	0.68	0.56

نتایج، داده‌های پیش‌بینی شده تراز آب دریاچه ارومیه برای سه ماه آتی نیز همانند ۲ ماه قبلی نشان داد مدل ترکیبی ROF-REPT دارای بالاترین عملکرد بوده و مدل منفرد REPT دارای قدرت پیش‌بینی کمتری می‌باشد (جدول ۷).

نتایج، داده‌های پیش‌بینی شده تراز آب دریاچه ارومیه برای دو ماه آتی با استفاده از مدل ROF-REPT، MAE، RMSE، NSE و RSR برابر با ۰/۲۶ متر، ۰/۱۹ متر، ۰/۷۶ و ۰/۵۲ دارای عملکرد بالاتری نسبت به مدل منفرد توسعه یافته دارد می‌باشد (جدول ۶).

Table 7- Model performance using quantitative criteria for  $WL_{(t+3)}$

جدول ۷- ارزیابی عملکرد مدلها با استفاده از معیارهای کمی برای  $WL_{(t+3)}$

	RMSE (m)	MAE (m)	NSE	RSR
ROF_REPT	0.34	0.26	0.53	0.69
REPT	0.37	0.29	0.45	0.74

نتایج نشان داد که با افزایش مقیاس زمانی پیش‌بینی از ۱ به ۳ ماه، قدرت پیش‌بینی کاهش یافته و عملکرد مدل‌ها همراه با عدم قطعیت بالاتری است. با این حال و بر اساس نتایج، پیش‌بینی تا ۳ ماه آتی خوب و قابل قبول است، در حالی که بیش از ۳ ماه پیش‌بینی پیش‌رو احتمالاً دارای عدم قطعیت بالایی می‌باشد. دلیل آن شاید به دلیل تغییرات شدید در نوسانات تراز سطح آب دریاچه در سال‌های اخیر ناشی از تغییرات آب و هوایی و احداث تعداد بسیار زیادی سد بر روی رودخانه‌های تغذیه‌کننده دریاچه باشد، که باعث تغییر در روند تراز سطح آب دریاچه شده باشد.

همانطور که قبلاً گفته شد، انتخاب بهترین و مؤثرترین سناریوی ورودی تأثیر زیادی بر نتیجه مدل‌سازی دارد. اگرچه بسیاری از محققین از روش‌های خودکار مانند تحلیل مؤلفه‌های اصلی ( $PCA^{13}$ ) (Barzegar et al., 2016)، یا آزمون گاما (Chobin et al., 2016) استفاده کردند، اما (Khosravi et al., 2020) نشان دادند که با استفاده از روش آزمون و خطا و ساخت چندین روش ترکیب ورودی مختلف، سناریوی ورودی بهتری با قدرت پیش‌بینی بالاتری ساخته می‌شود. دلیل آن این است که آن روش‌ها فقط چندین متغیر ورودی محدود را برای ساخت مدل ورودی در نظر می‌گیرند که در بیشتر موارد دارای قدرت پیش‌بینی پایین‌تری هستند. در پژوهش حاضر، بهترین ترکیب ورودی از متغیرها شامل داده‌های تراز آب دریاچه است، در حالی که متغیرهای هواشناسی نه به تنهایی و نه با ترکیب با سایر متغیرها قادر به پیش‌بینی صحیحی از تراز آب دریاچه نیستند. به دلیل اینکه دبی ورودی به دریاچه طی سال‌های اخیر به دلیل سدسازی‌های بی‌رویه به شدت دست‌خوش تغییر شده و از روندی که در درازمدت داشته است خارج شده، بنابراین متغیر دبی ورودی دارای پتانسیل کافی نمی‌باشد. از طرف دیگر، به دلیل تغییر در دبی ورودی، تسریع روند کاهش تراز آب دریاچه و تغییر شدید در تخریب از سطح دریاچه، این متغیرها دارای پتانسیل کافی برای برآورد تراز آب دریاچه ارومیه نیستند.

مزایای مدل REPT را می‌توان به صورت زیر خلاصه کرد:

- ۱- برآزش بیش از حد را در درخت‌های تصمیم کاهش می‌دهد،
- ۲- به بهبود دقت مدل‌سازی کمک می‌کند،
- ۳- برای مدل‌سازی‌های مکانی و رگرسیون انعطاف‌پذیر است،
- ۴- با مقادیر طبقه‌بندی شده و پیوسته به خوبی کار می‌کند،
- ۵- نسبت به مقادیر گم‌شده موجود در داده‌ها حساسیت پایینی دارد،

- ۶- چون از تکنیک رول بیس (یا بر پایه قاعده و قانون) بهره می‌گیرد، در نتیجه به نرمال‌سازی داده‌ها نیازی ندارد،
- ۷- همچنین، با هرس کردن مدل، مدل ساده‌تری ساخته که هیچ تأثیری بر روی افت و دقت مدل‌سازی ندارد.

Ehteram et al. (2021) مدل‌های ANN<sup>14</sup> و ANFIS<sup>15</sup> بهینه شده با استفاده از الگوریتم بهینه‌سازی آفتابگردان (SO<sup>16</sup>) را برای پیش‌بینی تراز سطح آب دریاچه ارومیه اعمال کردند. آنها بیان کردند که بهترین عملکرد مدل با استفاده از سه روز تأخیر در داده‌های بارندگی و دما به عنوان ورودی به دست می‌آید. در تحقیق حاضر بیان شد که ترکیب ورودی مختلف با استفاده از بارندگی، جریان و تبخیر آزمایش شده و این ورودی‌ها قادر به پیش‌بینی دقیق تراز سطح آب دریاچه ارومیه نیستند. Ehteram et al. (2021) بیان کردند که بهترین عملکرد آنها دارای  $RMSE=1.92$  متر برای مرحله اعتبارسنجی است در حالی که  $RMSE$  تحقیق حاضر در بدترین شرایط (یعنی پیش‌بینی برای سه ماه آینده) با ضعیف‌ترین مدل  $0.34$  متر است. نتیجه ضعیف آنها سه دلیل دارد، اولی در انتخاب اشتباه متغیرهای آب و هواشناسی به عنوان ورودی مدل، در حالی که آنها قادر به پیش‌بینی دقیق نیستند، دلیل دوم ساختار مدل است، همانطور که مدل‌های داده‌کاو پیشرفته در تحقیق حاضر قدرت پیش‌بینی بالاتری را نسبت به مدل‌های یادگیری ماشین سنتی دارند و همچنین طول داده‌های مدل‌سازی و اعتبارسنجی متفاوت می‌باشد.

(Shiri et al., 2016) چندین مدل سنتی (ELM<sup>17</sup>، ANN و GP<sup>18</sup>) را برای پیش‌بینی تراز سطح آب دریاچه ارومیه برای ۱ و ۷ روز آینده اعمال کردند و بیان کردند که  $RMSE$  آنها برای هر مدل توسعه‌یافته نزدیک به  $0.3$  است. اگرچه آنها به دلیل استفاده از داده‌های روزانه و مجموعه داده‌های طولانی‌تر برای مدل‌سازی، به نتیجه بهتری رسیدند، اما نتیجه آنها چندان کاربردی نیست، زیرا نوسان سطح آب دریاچه برای ۱ یا ۷ روز آینده چندان مهم نیست، در حالی که بهتر است مدلی توسعه داده شود که بتواند تراز سطح آب دریاچه‌ها را برای چندین ماه یا سال آینده پیش‌بینی کند. همچنین، آنها مجموعه داده‌های روزانه را از سال‌های ۱۹۷۲ تا ۲۰۰۳ اعمال کردند، زمانی که تغییرات و نوسانات سطح آب دریاچه ناچیز بود. در حالی که اخیراً تراز سطح آب دریاچه ارومیه به شدت کاهش یافته است و این باعث می‌شود که روندها تغییر یافته و قدرت پیش‌بینی مدل‌ها کاهش یابد.

## پی‌نوشت‌ها

- 1- Rapid
- 2- Machine Learning
- 3- Root Mean Square Error
- 4- Molecular Beam Electroplating
- 5- Nash-Sutcliffe Efficiency
- 6- Support Vector Regression
- 7- Grey Wolf Optimizer
- 8- Additive Regression
- 9- Rotation Forest
- 10- Base Learning Algorithm
- 11- Bootstrap
- 12- Base Classifier
- 13- Principale Component Analysis
- 14- Artificial Neural Network
- 15- Adaptive Neuro Fuzzy Inference System
- 16- Sunflower Optimization algorithm
- 17- Ensemble Learning Machine
- 18- Genetic Programming

در خصوص عدم قطعیت‌های موجود در تحقیق حاضر می‌توان به عدم قطعیت در داده‌ها و کیفیت آن‌ها، عدم قطعیت مربوط به انتخاب ایستگاه مناسب برای مدل‌سازی، عدم قطعیت مربوط به نمونه برداری برای ساختن مدل یادگیری ماشین، عدم قطعیت ناشی از تقسیم داده‌ها برای آموزش و اعتبارسنجی به صورت تصادفی، انتخاب مدل مناسب یادگیری ماشین و در نهایت ساخت مدل بهینه اشاره کرد.

## ۵- نتیجه‌گیری

در این تحقیق، تراز سطح آب دریاچه ارومیه در ۱، ۲، و ۳ ماه آینده با استفاده از ۲ مدل شامل پس هرس کردن درخت به روش REPT، و مدل ترکیبی REPT با مدل جنگل چرخان (ROF-REPT) توسعه و پیش‌بینی شد. داده‌های سری زمانی سطح آب از سال ۲۰۰۱ تا ۲۰۲۰ گردآوری و برای مدل‌سازی استفاده شد. همچنین، سناریوهای مختلف ورودی برای یافتن مؤثرترین سناریو ورودی ساخته و آزمایش شدند. در نهایت مدل‌های توسعه‌یافته از طریق چندین معیار بصری و کمی ارزیابی شدند. نتایج به دست آمده به شرح زیر است:

۱- مدل‌های توسعه یافته قادر به پیش‌بینی با دقت بالای تراز سطح آب دریاچه ارومیه تا ۳ ماه آینده هستند؛

۲- با افزایش مقیاس پیش‌بینی (یعنی از ۱ به ۳ ماه آینده)، دقت پیش‌بینی کاهش می‌یابد و مدل دارای عدم قطعیت بالا و قدرت پیش‌بینی پایین‌تری می‌شود؛

۳- مدل ترکیبی، دقت مدل منفرد را برای تمامی گام‌های زمانی افزایش داده است؛

۴- بهترین سناریو ورودی تنها با استفاده از داده‌های تراز سطح آب قبلی قابل دست‌یابی است؛

۵- داده‌های هواشناسی-هیدرولوژی از قبیل بارندگی، تبخیر و جریان رودخانه قادر به پیش‌بینی دقیق سطح آب در منطقه مورد مطالعه نیستند.

## ۶- مراجع

- Ahmadinasab A, Bahmanesh J, Khalili K (2013) Forecasting the water level of Lake Urmia using the genetic programming method. The Second National Conference on Environment, Energy and Biodefense (In Persian)
- Bahmanesh J, Morouj M, Khalili K (2014) Forecasting the level of Lake Urmia using linear time series model. *Water and Soil Conservation Research* 22(5):287-296
- Hassanzadeh Y, Alami MT, Farzin S, Sheikhl-Islami S R, Hassanzadeh A (2018) Investigating the chaotic nature of daily water level fluctuations in Lake Urmia. *Journal of Civil and Environmental Engineering, Tabriz University* 42(1):127-139 (In Persian)
- Karimzadegan S, Bahmanesh J, Rezaei H (2017) The application of Bayesian network and least squares vector machine methods in predicting the water level of Lake Urmia. *Water and Soil Conservation Research (Agricultural Sciences and Natural Resources)* 25(3):193-207
- Kamasi M, Nowzari H, Qashlaghi N (2015) Water level forecasting in Lake Urmia using time series methods, artificial neural network and neural-wavelet network. *Irrigation and Water Engineering of Iran* 6(24):64-77
- Malkani L, Khaleghi S (2016) Forecasting the water level of Lake Urmia using fuzzy model. *International Conference on Geographical Consequences and Environmental Effects of Urmia Lake Conditions, Tabriz University* 528(396):785-801
- Mirchi A, Azarde M, Nazemi A, Nasrollahi N, Hasanzadeh E (2015) Aral Sea syndrome desiccates Lake Urmia: Call for action. *Journal of Great Lakes Research* 41(1):307-311
- Amasyalı MF (2008) Yeni makine öğrenmesi metotları ve ilaç tasarımına uygulamaları. Thesis (Ph.d.), Yıldız of Technical University. Ayerdi B & Graña M (2014) Hybrid extreme rotation forest, *Neural Networks* 5(52):33-42
- Barzegar R, Aalami MT, Adamowski J (2020) Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment* 34(18):415-433.
- Barzegar R, Fijani E, Moghaddam AA, Tziritis E (2016) Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. *Science of the Total Environment* 599-600(12):20-31
- Bui DT, Ho T C, Pradhan B, Pham B T, Nhu V H, Revhaug I (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences* 75(8):110-121
- Bui D T, Pradhan B, Lofman O, Revhaug I, & Dick O B (2012) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Computers & Geosciences* 45(29):199-211
- Choubin B, Zehtabian GR, Azareh A, Rafiei-Sardooi E, Sajedi-Hosseini F, Kisi O (2018) Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. *Environmental Earth Sciences* 77(8):1-13
- Ehteram M, Ferdowsi A, Faramarzpour M, Mohammed Sami Al-Janabi A, Al-ansari N, Dhanraj Bokde N, Yaseen ZM (2021) Hybridization of artificial intelligence models with nature inspired optimization algorithms for Lake water level prediction and uncertainty analysis. *Alexandria Engineering Journal* 60(52):2193-2208
- Freund Y, and L Mason (1999) The alternating decision tree-learning algorithm. In *Proc 16th International Conf on Machine Learning, San Francisco, CA, USA, (1999)*. Morgan Kaufmann Publishers Inc 116(36):124-133
- Fry LM, Apps D, & Gronewold A D (2020) Operational seasonal water supply and water level forecasting for the Laurentian great lakes. *Journal of Water Resources Planning and Management* 146(9):225-236
- Huang A, Rao YR, Lu Y, Zhao J (2010) Hydrodynamic modeling of Lake Ontario: An intercomparison of three models. *Journal of Geophysical Research* 115(12):C12076
- Khatibi R, Ghorbani MA, Naghipour L, Jothiprakash V, Fathima TA, Fazelifard MH (2014) Inter-comparison of time series models of lake levels predicted by several modeling strategies. *Journal of Hydrology* 511(98):530-545
- Khosravi K, Safari M J, Sheikh Khozani Z, Crookston B, Golkarian A (2022) Stacking ensemble-based hybrid algorithms for discharge computation in sharp-crested labyrinth weirs. *Application of soft computing* 396(125):160-172

- Khosravi K, Barzegar R, Miraki S, Adamowski J, Daggupati P (2020) Stochastic modeling of groundwater fluoride contamination: Introducing lazy learners. *Groundwater* 58(5):723-734
- Lin P, Yang Z L, CAI X, & David C H (2015) Development and evaluation of a physically based lake level model for water resource management: A case study for Lake Buchanan, Texas. *Journal of Hydrology: Regional Studies* 4(11):661-674
- Mohammadi B, Guan Y, Aghelpour P, Emamgholizadeh S, Zola RP, Zhang D (2020) Simulation of Titicaca Lake water level fluctuations using hybrid machine learning technique integrated with Grey Wolf Optimizer Algorithm. *Water* 12(11):315-329
- Mohebzadeh H (2018) Extracting A-L relationship for Urmia Lake, Iran Using MODIS NDVI/NDWI Indices. *Journal of Hydrogeology and Hydrologic Engineering* 78(62):115-124
- Ozcift A & Gulten A (2011) Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine* 104(3):443-451
- Parsinejad M, Rosenberg D, Alizadeh Y, Ghale G, et al. (2022) 40-years of Lake Urmia restoration research: Review, synthesis and next steps. *Science of the Total Environment* 832(520):668-679
- Pérez-Chacón R, Asencio-Cortés G, Martínez-Álvarez F, & Troncoso A (2020) Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand. *Information Sciences* 540(122):160-174
- Quilty J, Adamowski J (2020) A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes. *Environmental Modelling and Software* 130(25):104-120
- Razmkhah A, Alvankar SR, Kakahaji A (2016) Modeling Lake Urmia water-level changes using local linear neuro-fuzzy method. *Journal of Water Sciences Research* 9(25):65-78
- Rodriguez JJ, Kuncheva LI, & Alonso CJ (2006) Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(10):1619-1630
- Shiri J, Shamshirband S, Kisi O, Karimi S, Bateni S M, Nezhad SHH (2016) Prediction of water level in the Urmia Lake using the extreme learning machine approach. *Water Resources Management* 30(14):5217-5229
- Tiwari MK, Chatterjee C (2011) A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *Journal of Hydroinformatics* 13(3):500-519
- Vaheddoost B, Aksoy H (2021) Regressive-stochastic models for predicting water level in Lake Urmia. *Hydrological Sciences Journal* 66(13):285-297
- Zhu S, Hrnjica B, Ptak M, Choiński A, Sivakumar B (2020) Forecasting of water level in multiple temperate lakes using machine-learning models. *Journal of Hydrology* 585(320):685-700
- Zhu S, Lu H, Ptak M, Dai J, Ji Q (2020) Lake water-level fluctuation forecasting using machine learning models: A systematic review. *Environmental Science and Pollution Research* 27(14):89-105