

## مدل‌سازی بارش - رواناب حوضه آجی‌چای با استفاده از مدل‌های جنگل تصادفی و شبکه عصبی مصنوعی

زینب بیگدلی<sup>۱\*</sup>، ابوالفضل مجنونی هریس<sup>۲</sup>، رضا دلیر حسن نیا<sup>۲</sup>، سپیده کریمی<sup>۳</sup>

تاریخ ارسال: ۱۴۰۱/۰۶/۳۱

تاریخ پذیرش: ۱۴۰۱/۱۲/۲۴

DOI: 10.22103/nrswe.2023.20278.1013

### چکیده

برآورد رواناب ناشی از وقوع بارندگی، گامی بسیار مهم در برنامه‌ریزی منابع آب به ویژه در آبخیزهای فاقد ایستگاه‌های هیدرومتری است. در این مطالعه به شبیه‌سازی بارش - رواناب ایستگاه آخولا واقع در حوضه آجی‌چای پرداخته شد و با استفاده از روش‌های داده‌کاوی و مقایسه عملکرد آنها، مناسب‌ترین مدل بارش - رواناب ارائه گردید. برای این منظور داده‌های مورد نظر (بارش، دبی، دما) بصورت ماهانه از سازمان‌های آب و هواشناسی استان‌های آذربایجان شرقی و غربی دریافت گردید. جهت شبیه‌سازی از مدل‌های داده‌کاوی جنگل تصادفی و شبکه عصبی مصنوعی استفاده گردید. مقایسه مقادیر رواناب ماهانه مشاهداتی با رواناب ماهانه تخمین زده شده توسط مدل‌ها با استفاده از معیارهای ارزیابی انجام شد. در این مطالعه مقادیر CC (ضریب همبستگی) برای مجموعه‌های تست در مدل جنگل تصادفی و شبکه عصبی مصنوعی به ترتیب برابر با ۰/۸۲ و ۰/۸۶ تعیین گردید. تحلیل نتایج نشان داد که برای ایستگاه آخولا مدل ANN عملکرد و کارایی بالاتری نسبت به مدل RF دارد. از نتایج دیگر این پژوهش می‌توان به سری زمانی بارش و رواناب ایستگاه طی ۲۰ سال اخیر اشاره کرد. با توجه به تحلیل روند من-کندال، در طی این ۲۰ سال روند مشخصی برای بارش بر روی حوضه آجی‌چای دیده نشد و نمودارهای سری زمانی نشان داد که بارش در این مناطق بصورت نوسانی بوده است. اما سری زمانی برای آجی‌چای در ایستگاه آخولا، نشان داد که روند کاملاً نزولی برای جریان آب رودخانه ثبت شده است که در واقع دلیل کاهش دبی ورودی به دریاچه ارومیه و پایین آمدن تراز آب دریاچه می‌باشد.

**واژگان کلیدی:** بارش - رواناب، جنگل تصادفی، حوضه آجی‌چای، مدل‌سازی، یادگیری ماشین.

\*۱ - نویسنده مسئول و دانشجوی دکتری، گروه مهندسی آب، دانشکده کشاورزی دانشگاه تبریز، تبریز، ایران، ایمیل: z.bigdeli.20@gmail.com

۲ - دانشیار گروه مهندسی آب، دانشکده کشاورزی دانشگاه تبریز، تبریز، ایران.

۳ - دانشجوی پست دکتری، گروه مهندسی آب، دانشکده کشاورزی دانشگاه تبریز، تبریز، ایران.

## مقدمه

به طور کلی بخش قابل توجهی از نزولات آسمانی با توجه به ویژگی‌های فیزیوگرافی آبخیزها به رواناب تبدیل می‌شود. رواناب حاصل از بارش و ذوب برف یکی از منابع اصلی برای تحقق نیازهای کشاورزی، صنعتی و مصارف خانگی است؛ به طوری که برنامه‌ریزان منابع آب با تکیه بر داده‌های جریان رواناب، تخصیص منابع آب به این مصارف را تعیین می‌کنند (Worland et al. 2018). بنابراین برآورد رواناب ناشی از وقوع بارندگی، گامی بسیار مهم در برنامه‌ریزی منابع آب است. تعداد زیادی از آبخیزها در ایران و بیشتر کشورهای در حال توسعه، فاقد ایستگاه‌های هیدرومتری هستند و یا داده‌های اندازه‌گیری شده از رواناب در آنها بسیار محدود است. به طوری که مناسب برای اهداف برنامه‌ریزی منابع آبی نیستند (Dastorani et al. 2018). لذا مدیران منابع آب به منظور برنامه‌ریزی به مدل‌های مختلف پیش‌بینی جریان رواناب وابسته هستند (Razavi and Coulibaly 2013; Luce 2014). برآورد رواناب تولید شده در هر حوضه، با دقت قابل قبول، بخشی مهم و جداناپذیر از اطلاعات لازم برای سیاست‌گذاری و مدیریت حوضه آبخیز می‌باشد. اطلاعات حاصل از مدل‌سازی بارش-رواناب و تخمین جریان آب در رودخانه‌ها در کارهای عمرانی، ساماندهی رودخانه‌ها، طراحی و برنامه‌ریزی منابع آب‌های سطحی و زیرزمینی و سامانه‌های هشدار سیل از اهمیت فراوانی برخوردار است. در این راستا، در دسترس بودن اطلاعات پیوسته بارش و سایر متغیرهای اقلیمی که می‌توانند برای مدل‌سازی مقدار رواناب مورد استفاده قرار گیرد بسیار ضروری می‌باشد (Lorrai and Sechi 1995). از سوی دیگر با توجه به محدودیت منابع آب شیرین و قابل استحصال، پیش‌بینی هرچه دقیق‌تر تغییرات دبی جریان در طول سال از ارکان اساسی برنامه‌ریزی و مدیریت منابع آب‌های سطحی می‌باشد. در هر حال در طرح‌های آبی لازم است که به نحوی از میزان تبدیل بارش به رواناب آگاه شد. در این راستا مدل‌های زیادی از اوایل دهه ۱۹۶۰ توسعه داده شده است تا فرایند بارش-رواناب را شبیه‌سازی کنند. مدل ابزاری است

که می‌تواند بخشی از واقعیت را شبیه‌سازی کند. هدف از مدل‌سازی، نشان دادن عملکرد و رفتار یک پدیده طبیعی تحت شرایط معین می‌باشد (Phomcha et al. 2011). به طور کلی، مدل‌های ریاضی شناخته شده بارش-رواناب را میتوان در دو گروه مدل‌های دارای مبنای فیزیکی (همچون مدل‌های هیدرولوژیکی مفهومی) و مدل‌های داده محور (همچون شبکه عصبی مصنوعی و ..) تقسیم بندی نمود. با استفاده از مدل‌های هوشمند می‌توان روابط ذاتی میان داده‌ها را که ریاضیات کلاسیک قادر به حل آنها نمی‌باشد، مدل‌سازی نمود. از جمله مدل‌های هوشمند می‌توان به مدل‌های برنامه‌ریزی بیان ژن (GEP)<sup>۱</sup>، ماشین بردار پشتیبان (SVM)<sup>۲</sup>، شبکه‌های عصبی مصنوعی (ANNs)<sup>۳</sup> و مدل‌های جدیدی همچون جنگل تصادفی (RF)<sup>۴</sup> اشاره کرد. تاکنون در مطالعات زیادی از مدل‌های هوشمند برای پیش‌بینی جریان رودخانه استفاده شده است که در ادامه به برخی از آنها اشاره می‌شود.

در داخل کشور (Hosseini 1999) از مدل شبکه عصبی مصنوعی برای پیش‌بینی هیدروگراف سیل در حوزه آبخیز باروت آجی، ناورود آسالم و ابوالعباس استفاده نموده است. نتایجی که از ارزیابی آبخیزها به دست آمد نشان می‌دهد که شبکه عصبی مصنوعی و شبکه عصبی-فازی تطبیقی در شرایط مختلف و با ترکیب‌های مختلف پارامترهای ورودی، نتایج متفاوتی از خود نشان می‌دهند، ولی در کل این دو روش به میزان قابل قبولی قادر به تخمین رواناب حاصل از بارش با به کارگیری پارامترهای ورودی مناسب و استفاده از ساختارهای مناسب شبکه عصبی مصنوعی<sup>۵</sup> و شبکه عصبی-فازی تطبیقی هستند. در مطالعه دیگر (Ebrahimi et al. 2011) در مدل‌سازی تغییرات دبی ماهانه رودخانه قره‌سو با استفاده از سری‌های زمانی و شبکه‌های عصبی مصنوعی و با به کارگیری داده‌های بارندگی، دما، تبخیر و دبی به این نتیجه رسیدند که نتایج حاصل از پیش‌بینی بر اساس مدل شبکه عصبی مصنوعی انطباق بیشتر با مقادیر مشاهده شده دارد. (Nourani et al. 2009) برای پیش‌بینی رواناب ماهانه و روزانه در حوضه آبخیز ليقوان چای از سه

<sup>۱</sup>-GeneXproTools<sup>۲</sup>-Support Vector Machine<sup>۳</sup>-Artificial Neural Network<sup>۴</sup>-Random Forest

مدل هوشمند (شبکه‌های عصبی مصنوعی، سیستم استنتاج فازی و سیستم عصبی-فازی تطبیقی) استفاده کردند و نتایج سه مدل مذکور را با نتایج به دست آمده از روش‌های رگرسیون خطی و مدل سری زمانی ARIMA مقایسه کردند که مدل‌سازی فازی (تطبیقی و استنتاجی) بیشترین مقدار نش-ساتکلیف و کمترین خطا را در پی داشت. (Poursalehi et al. 2022) عملکرد الگوریتم جنگل تصادفی در پیش‌بینی نوسانات سطح ایستابی در مقایسه با دو مدل درخت تصمیم و شبکه عصبی مصنوعی (مطالعه موردی: آبخوان آزاد دشت بیرجند) را مطالعه کردند. نتایج حاصل از شبیه‌سازی با استفاده از الگوریتم جنگل تصادفی نشان داد که این الگوریتم توانایی نسبتاً زیادی در شبیه‌سازی تراز سطح ایستابی آبخوان دارد. در جدیدترین مطالعه انجام گرفته توسط (Zarei et al. 2022) پتانسیل وقوع سیل با استفاده از مدل‌های داده‌کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی (مطالعه موردی: حوزه آبخیز فریزی) را ارزیابی نمودند. بر اساس نتایج، هم در داده‌های تعلیمی و هم در داده‌های آموزشی، الگوریتم (SVM)<sup>۶</sup> و مدل الگوریتم نوین جنگل تصادفی دارای نتایج دقیق‌تری می‌باشد.

در خارج از ایران (Nayak et al. 2004) از شبکه عصبی-فازی تطبیقی برای مدل‌سازی سری‌های زمانی هیدرولوژیکی در رودخانه بیترانی در اندونزی استفاده کردند. این مدل در مقایسه با سایر روش‌های مدل‌سازی از قبیل شبکه عصبی مصنوعی و شبکه عصبی-فازی تطبیقی با توجه به نمایه‌های ارزیابی، جریان مشاهداتی را بهتر برآورد کرد. (Nilsson et al. 2005) با استفاده از مدل‌های مفهومی و شبکه‌های عصبی مصنوعی و ترکیب آنها، مقادیر ماهانه رواناب خروجی از دو حوزه آبخیز در نروژ را شبیه‌سازی نمودند و به این نتیجه رسیدند که روش شبکه عصبی مصنوعی و ترکیب آن با مدل‌های مفهومی برآورد بهتری از رواناب ماهانه را برای هر دو حوزه آبخیز داده است. (Firat and Gungor 2006) شبکه عصبی-فازی تطبیقی را برای پیش‌بینی جریان رودخانه بزرگ مئاندر در ترکیه مورد ارزیابی قرار داده و به این نتیجه رسیدند که شبکه عصبی-

فازی تطبیقی می‌تواند با موفقیت، صحت و قابلیت اطمینان بالا برای پیش‌بینی جریان رودخانه به کار رود. (Kisi 2010) به منظور پیش‌بینی رواناب روزانه در دو ایستگاه کربوک و درک ویران روی رودخانه فلیوس واقع در کشور ترکیه از رگرسیون مویک استفاده نمود و نتایج آن را با شبکه عصبی مصنوعی و سری زمانی مقایسه کرد. نتایج نشان داد که رگرسیون مویک با دقت بالاتری نسبت به سری زمانی و شبکه عصبی مصنوعی، رواناب را پیش‌بینی نموده است. (Kurtulus and Razack 2010) دو روش شبکه عصبی مصنوعی و شبکه عصبی-فازی تطبیقی (ANFIS) را برای پیش‌بینی دبی روزانه در یک رودخانه کارستی در جنوب شرقی فرانسه به کار بردند و نتیجه گرفتند که ANFIS به طور کلی جواب مناسب‌تری نسبت به شبکه عصبی مصنوعی ارائه می‌نماید. (Toker and Markus 2000) با استفاده از شبکه عصبی مصنوعی رابطه بارندگی-رواناب را برای پیش-بینی جریان ماهانه در سه حوزه آبخیز با شرایط آب و هوایی و فیزیوگرافی متفاوت در مرینلد آمریکا مدل‌سازی کرده و نتایج حاصله را با مدل مفهومی Watbal مقایسه کردند. نتایج نشان داد مدل شبکه عصبی مصنوعی در همه موارد دارای کارایی بالاتری است. (Pham et al. 2021) برای پیش‌بینی جریان ۸۶ حوضه آبخیز واقع در شمال غربی اقیانوس آرام، از الگوریتم جنگل تصادفی (RF) استفاده کردند و عملکرد این الگوریتم را با مدل‌های رگرسیون خطی چندگانه (MLR)<sup>۷</sup> و الگوریتم Naïve مورد مقایسه قرار دادند. نتایج نشان داد که بیانگر عملکرد مدل RF در حوضه‌های آبخیز با رژیم برفی بهتر از دو مدل دیگر بوده است. (Hussain et al. 2020) جهت پیش‌بینی جریان رودخانه هنزه در پاکستان از مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)<sup>۸</sup>، رگرسیون بردار پشتیبان (SVR)<sup>۹</sup> و جنگل تصادفی (RF)<sup>۱۰</sup> استفاده نموده اند. مجموعه‌ای از پنج پارامتر به اضافه سری زمانی جریان رودخانه با تاخیرهای مختلف مورد بررسی قرار گرفتند و در نتیجه مدل‌های MLP، RF و SCR بالاترین عملکرد را در پیش‌بینی جریان رودخانه هنزه داشتند. همچنین نتایج

<sup>۶</sup>- Support Vector Machine

<sup>۷</sup>- Multiple Linear Regression

<sup>۸</sup>-Multi-Layer Perceptron

<sup>۹</sup>-Support Vector Regression

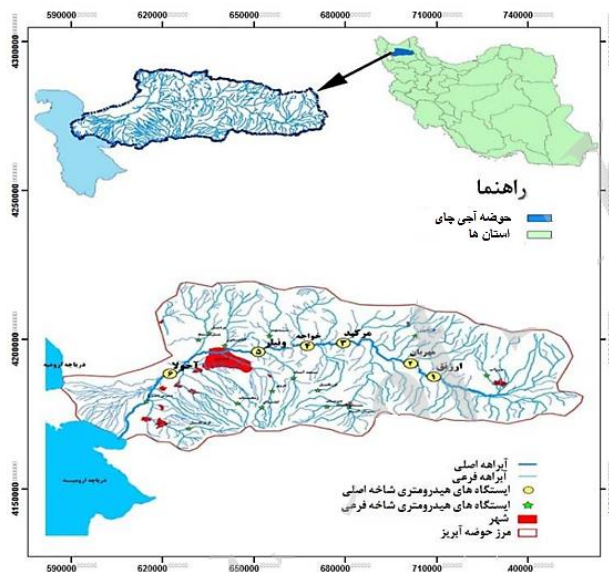
<sup>۱۰</sup>- Random Forest

حوضه بشمار می‌آیند. این حوضه در محدوده  $30^{\circ}45'$  تا  $30^{\circ}48'$  عرض شمالی واقع شده است که موقعیت آن در شکل (۱) ارائه شده است. حوضه آبریز آبی‌چای از شمال به حوضه آبریز اهرچای، از جنوب به حوضه آبریز شهرچای، از شرق به حوضه آبریز بالخواجای و از غرب به دریاچه ارومیه محدود می‌شود. رودخانه آبی‌چای در دره‌ای که ما بین کوه‌های ارسباران، قوشه‌داغ و سبلان از شمال و کوه‌های بزقوش داغ و سهند از جنوب قرار گرفته است در جهت عمومی شرقی-غربی جریان می‌یابد. از ابتدای زیرحوضه آبی‌چای، رودخانه‌های بیوک‌چای، تاجیارچای، رازلیق‌چای و ونق‌چای، اوجان‌چای و در دشت تبریز مهران رود و گمناب‌چای و سینه‌چای از سمت راست و رودخانه‌های سردرود و عنصررود به آبی‌چای پیوستند و در انتهای دشت به دلیل شیب بسیار کم سطح زمین به چند شاخه تقسیم می‌شود تا در نهایت به دریاچه ارومیه تخلیه شود. میانگین سالانه دمای متوسط روزانه در حوضه مطالعاتی از حدود  $10^{\circ}$  درجه در مناطق اطراف دریاچه ارومیه تا کمتر از  $2/5^{\circ}$  درجه در سال در ارتفاعات سهند و سبلان متغیر است.

نشان داد که مدل‌های مبتنی بر الگوریتم‌های یادگیری ماشینی به‌ویژه مدل RF، می‌توانند با دقت بسیار بالایی در پیش‌بینی جریان رودخانه به کار گرفته شوند. همچنین در زمینه استفاده از مدل‌های هوشمند برای پیش‌بینی جریان می‌توان به مطالعات (Hassanzadeh et al. 2015; Adnan et al. 2018; Shafeizadeh et al. 2019) اشاره کرد. نوآوری این مطالعه شبیه‌سازی فرایند بارش-رواناب در زیرحوضه رودخانه آبی‌چای (با ۶ ایستگاه) در استان آذربایجان شرقی با استفاده از نرم افزار STATISTICA برای ایجاد مدل بارش-رواناب و اجرای مدل‌های جنگل تصادفی و شبکه عصبی مصنوعی (MLP 2-5-1) می‌باشد. هدف از این مطالعه تعیین مدل برتر برای شبیه‌سازی بارش-رواناب در ایستگاه نهایی زیرحوضه رودخانه آبی‌چای (آخولا) می‌باشد. در واقع سؤال اساسی این تحقیق این است که آیا برای شبیه‌سازی فرایند بارش-رواناب در این زیرحوضه، در شرایط کمبود داده و اطلاعات، مدل جنگل تصادفی و شبکه عصبی مصنوعی (MLP 2-5-1) می‌تواند ابزار مناسبی باشد یا خیر!

## مواد و روش‌ها

رودخانه آبی‌چای در ناحیه شمال غربی ایران در استان آذربایجان شرقی قرار گرفته و شهرهای تبریز، آذرشهر، سراب، بستان‌آباد، هریس و اسکو نقاط مهم شهری این



شکل (۱): موقعیت جغرافیایی زیرحوضه آبی‌چای

ایستگاه منتهی به دریاچه ارومیه) در مطالعه مدل‌سازی شدند. در جدول (۱) خلاصه‌ای از مشخصات آماری ایستگاه‌های مورد مطالعه ارائه شده است. نکته‌ای که در این جدول قابل ذکر است، رودخانه‌های زیرحوضه آبی‌چای می‌باشد که تمام این رودخانه‌های فصلی یا دائمی به شاخه اصلی رودخانه آبی‌چای می‌ریزند و جریان رودخانه آبی‌چای در نهایت در ایستگاه آخولا ثبت می‌گردد. در این مطالعه بارش روی حوضه آبی‌چای و رواناب حاصله از آبی‌چای در ایستگاه آخولا برای مدل‌سازی بارش - رواناب استفاده شده است.

داده‌های مورد استفاده برای مدل‌سازی در این پژوهش، شامل داده‌های جریان رودخانه‌های ورودی به رودخانه آبی‌چای (ماهانه)، باران (ماهانه)، دمای میانگین (ماهانه) و دبی (ماهانه) در یک دوره ۲۰ ساله (از سال ۱۳۷۷ تا ۱۳۹۷) می‌باشد. این اطلاعات از ۶ ایستگاه سینوپتیک و هیدرومتری (ارزنق، مهربان، مرکید، خواجه، و نیار و آخولا) در حوضه آبی‌چای از سازمان آب استان و هواشناسی آذربایجان شرقی تهیه گردید. برای مدل‌سازی بارش-رواناب، ابتدا بارش متوسط بر روی حوضه آبی‌چای محاسبه شد و سپس به همراه دبی خروجی از ایستگاه آخولا (آخرین

جدول(۱): مشخصات آماری ایستگاه‌های مورد مطالعه

زیرحوضه	مساحت (m <sup>3</sup> )	ایستگاه‌های منتخب	نام رودخانه	ارتفاع ایستگاه	متوسط بارش (mm)	متوسط دما (C°)
بیلوردی - دوز دوز	۲۳۷۱	ارزنق	آبی‌چای	۱۶۰۰	۲۵۳/۱۴	۹/۷۹
بیلوردی - دوز دوز	۲۳۷۱	مهربان	چکه‌چای	۱۵۱۰	۲۶۹/۳۲	۹/۷۹
بیلوردی - دوز دوز	۲۳۷۱	مرکید	آبی‌چای	۱۵۱۸	۲۳۲/۶۰	۱۲/۰۹
تبریز	۵۴۲۸	خواجه	پاژچای	۱۵۰۰	۲۹۶/۳۰	۱۲/۱
تبریز	۵۴۲۸	ونیار	آبی‌چای	۱۴۵۰	۲۰۴	۱۲/۸۲
تبریز	۵۴۲۸	آخولا	آبی‌چای	۱۳۱۰	۲۱۳/۱۲	۱۲/۸۲

بیان ساده، جنگل تصادفی چندین درخت تصمیم می‌سازد و آنها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری به دست آیند (Catani et al. 2013; Nicodemus et al. 2011; Youssef et al. 2015). این الگوریتم را نخستین بار (Breiman 2001) ایجاد و توسعه داد. مدل پیش‌بینی کننده RF بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوط استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهد (Catani et al. 2013). درختان تصادفی با در نظر گرفتن بردار ورودی، آن را با هر درخت در جنگل، طبقه‌بندی کرده و خروجی، کلاس‌هایی هستند که از اکثریت آرا دریافت شده است. در این مدل از دو عامل میانگین کاهش دقت و میانگین کاهش جینی برای تعیین اولویت تاثیر هر یک از عوامل موثر استفاده می‌شود (Nicodemus et al. 2011). الگوریتم جنگل تصادفی (RF) مبتنی بر دسته‌ای از درخت‌های تصمیم است و در حال

### الگوریتم جنگل تصادفی (Random-Forest)

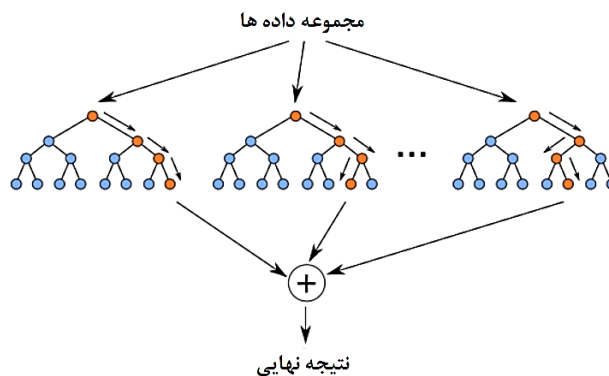
جنگل تصادفی، نوعی الگوریتم یادگیری ماشین و نظارت شده با قابلیت استفاده آسان است که اغلب اوقات نتایج بسیار خوبی را حتی بدون تنظیم پارامترهای آن، فراهم می‌کند (Catani et al. 2013). جنگل‌های تصادفی یک نوع مدرن از روش‌های درخت-پایه هستند که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی می‌باشند (Breiman 2001). این الگوریتم به دلیل سادگی و قابلیت استفاده، هم برای دسته‌بندی و هم رگرسیون، از پرکاربردترین الگوریتم‌های یادگیری ماشین محسوب می‌شود. همان‌طور که از نام آن مشهود است، این الگوریتم جنگلی را به‌طور تصادفی می‌سازد. جنگل ساخته شده در واقع گروهی از درخت‌های تصمیم<sup>۱۴</sup> است. کار ساخت جنگل با استفاده از درخت‌ها اغلب اوقات به روش کیسه‌گذاری<sup>۱۵</sup> انجام می‌شود. ایده اصلی روش کیسه‌گذاری آن است که ترکیبی از مدل‌های یادگیری، نتایج کلی مدل را افزایش می‌دهد. به

<sup>۱۴</sup> - Decision Trees

<sup>۱۵</sup> - Bagging

بستگی دارد: الف- همبستگی بین هر دو درخت در جنگل: افزایش همبستگی نرخ خطای جنگل را افزایش می‌دهد. ب- قدرت هر یک از درختان در جنگل: هر درخت با نرخ خطای کم، یک طبقه‌بند قوی است. افزایش قدرت هر یک از درختان نرخ خطای جنگل را کاهش می‌دهد. از جمله مزیت‌های جنگل تصادفی می‌توان به موارد زیر اشاره نمود: ۱. در میان الگوریتم‌های فعلی از نظر دقت کم نظیر است. ۲. روی داده‌های بسیار بزرگ قابل اجراست. ۳. می‌تواند هزاران متغیر را بدون حذف متغیرها مدیریت کند. ۴. برآوردی از مهم‌ترین متغیرها در طبقه‌بندی می‌دهد. ۵. راهکارهایی برای برآورد داده‌های گم شده دارد (Sohrabi et al. 2021).

حاضر یکی از بهترین الگوریتم‌های یادگیری است. هر درخت به صورت زیر تشکیل می‌شود: اگر  $N$  تعداد حالت‌ها در مجموعه داده‌های train (مجموعه کار) باشد،  $N$  حالت به صورت تصادفی با جای‌گذاری از داده‌های اصلی، نمونه‌گیری می‌شود. این نمونه مجموعه کار برای این درخت می‌باشد. اگر  $M$  متغیر داشته باشیم و  $m$  را کوچکتر از  $M$  در نظر بگیریم، به طوری که در هر گره،  $m$  متغیر به صورت تصادفی از  $M$  انتخاب می‌شوند و بهترین جداسازی روی این  $m$  متغیر برای جداسازی گره استفاده می‌شود. مقدار  $m$  در طول ساخت جنگل ثابت در نظر گرفته می‌شود. هر درخت به اندازه ممکن بزرگ می‌شود. هیچ هرسی وجود ندارد. نرخ خطای جنگل به دو مورد زیر



شکل (۲): ساختار الگوریتم جنگل تصادفی

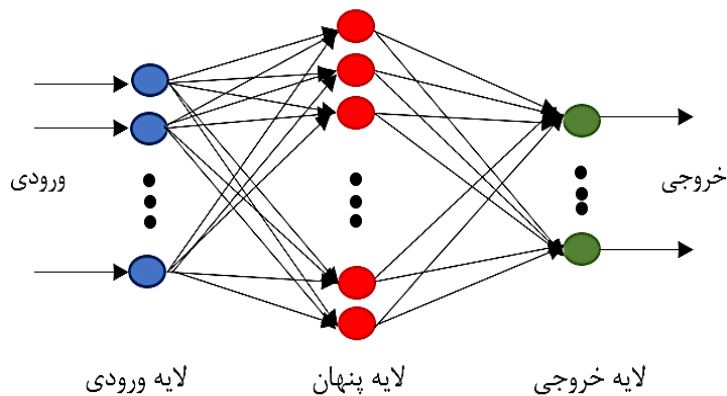
سیگموئید و تانژانت هیپربولیک می‌باشد. هر شبکه با دریافت مثال‌هایی آموزش می‌بیند. آموزش فرآیندی است که در نهایت منجر به یادگیری می‌شود. یادگیری شبکه، زمانی انجام می‌شود که وزن‌های ارتباطی بین لایه‌ها چنان تغییر کند که اختلاف بین مقادیر پیش‌بینی شده و محاسبه شده در حد قابل قبولی باشد (Dayhoff 1990). ساختار شبکه عصبی مصنوعی پرسپترون چند لایه (MLP) در شکل (۳) نشان داده شده است. شبکه پرسپترون چند لایه (MLP) نوعی از شبکه‌های عصبی مصنوعی است که بر مبنای یک واحد محاسباتی به نام پرسپترون ساخته می‌شود. یک پرسپترون، برداری از ورودی‌ها با مقادیر حقیقی را گرفته و یک ترکیب خطی از این ورودی‌ها را محاسبه می‌کند. اگر مقدار حاصل از یک مقدار آستانه بیشتر بود خروجی پرسپترون برابر با ۱ و در غیر این صورت معادل ۰ خواهد بود. این شبکه شامل یک لایه ورودی، یک

### شبکه عصبی مصنوعی (ANN)

شبکه عصبی مصنوعی (ANN)، یکی از روش‌های محاسباتی است که به کمک فرآیند یادگیری و با استفاده از پردازشگرهایی بنام نرون تلاش می‌کند با شناخت روابط ذاتی بین داده‌ها، نگاهی میان فضای ورودی (لایه ورودی) و فضای مطلوب (لایه خروجی) ارائه دهد. این سیستم از شمار زیادی عناصر پردازشی فوق‌العاده به هم پیوسته به نام نرون تشکیل شده که برای حل یک مسئله با هم هماهنگ عمل می‌کند. هر شبکه عصبی از سه لایه ورودی، مخفی و خروجی تشکیل شده است. در هر لایه تعدادی پردازشگر به نام نرون وجود دارد. لایه یا لایه‌های مخفی، اطلاعات دریافت شده از لایه ورودی را پردازش کرده و در اختیار لایه خروجی قرار می‌دهند. تابع انتقال بیان‌کننده پاسخ هر نرون به سیگنال ورودی آن نرون است. توابع انتقال معمول مورد استفاده در شبکه‌های عصبی مصنوعی تابع

لایه، ورودی با توجه به ساختار شبکه عصبی مصنوعی، ویژگی‌های عمده آن، سرعت بالای پردازش، توانایی یادگیری الگو به روش اراده الگو، توانایی تعمیم دانش پس از یادگیری، انعطاف پذیری در برابر خطاهای ناخواسته و عدم ایجاد اختلال قابل توجه در صورت بروز اشکال در بخشی از اتصال‌ها به دلیل توزیع وزن‌های شبکه است ( Khanna 1990).

یا چند لایه پنهان و یک لایه خروجی است. برای آموزش این شبکه، معمولاً از الگوریتم پس انتشار (BP) استفاده می‌شود. در طی آموزش شبکه MLP به کمک الگوریتم یادگیری BP، ابتدا محاسبات از ورودی شبکه به سوی خروجی شبکه انجام می‌شود و سپس مقادیر خطای محاسبه شده به لایه‌های قبل انتشار می‌یابد. در ابتدا، محاسبه خروجی به صورت لایه به لایه انجام می‌شود و خروجی هر



شکل (۳): ساختار شبکه عصبی پرسپترون چند لایه (Khademi et al. 2017)

نش- ساتکلیف می‌تواند از منفی بی‌نهایت تا ۱ تغییر کند. هرچه مقادیر به عدد یک نزدیک‌تر باشد، مدل برازش بهتری دارد و عدد یک برازش عالی را نشان می‌دهد. RMSE جذر میانگین مربع خطاهای داده‌های محاسباتی و مشاهداتی را بیان می‌کند. واضح است که هر چه مقدار این عدد کمتر باشد، عملکرد مدل در شبیه‌سازی داده‌ها بهتر است.

امروزه یکی از ابزارهای تعیین تغییرهای اقلیمی تحلیل روند می‌باشد. روش‌های متفاوتی برای تعیین روند وجود دارند که به دو دسته تقسیم می‌شوند: روش پارامتریک و ناپارامتریک. در اصل گفته می‌شود که آزمون‌های پارامتریک در صورتی که داده‌ها به صورت نرمال باشند، قوی‌تر از آزمون‌های ناپارامتریک هستند ولی شرایط داده‌های غیرنرمال آزمون ناپارامتریک قوی‌تر از پارامتریک است (Sabouhi and Soltani 2008). از آنجایی که توزیع تعدادی از سری‌های اقلیمی از جمله بارش و دبی نرمال نیست، در چنین حالت‌هایی روش من- کندال مناسب‌تر است. تحلیل روند در این مطالعه با استفاده از آزمون ناپارامتریک من-کندال انجام شد. آزمون من-کندال ابتدا توسط Mann (1945) مطرح و سپس توسط Kendall (1975) بسط و توسعه یافت (Serrano et al. 1999).

### معیارهای ارزیابی

برای ارزیابی مدل‌ها و مقایسه کارایی آنها، در این پژوهش از معیارهای جذر میانگین مربعات خطا (RMSE)، ضریب همبستگی (r)، ضریب تعیین (R<sup>2</sup>) و ضریب کارایی نش- ساتکلیف (NS) استفاده می‌شود. در ادامه روابط این معیارها ارائه شده است:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (1)$$

$$r = \frac{n(\sum_{i=1}^n O_i P_i) - (\sum_{i=1}^n O_i)(\sum_{i=1}^n P_i)}{\sqrt{(n \sum_{i=1}^n O_i^2 - (\sum_{i=1}^n O_i)^2)(n \sum_{i=1}^n P_i^2 - (\sum_{i=1}^n P_i)^2)}} \quad (2)$$

$$R^2 = \frac{[\sum_{i=1}^n (O_i - \bar{O}_i) \cdot (P_i - \bar{P}_i)]^2}{\sum_{i=1}^n (O_i - \bar{O}_i) \cdot \sum_{i=1}^n (P_i - \bar{P}_i)} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - \bar{P})^2} \quad (4)$$

در روابط فوق، n تعداد داده‌ها، P<sub>i</sub> داده‌های مشاهداتی، Q<sub>i</sub> میانگین داده‌های مشاهداتی و O<sub>i</sub> داده‌های محاسباتی می‌باشد. هر چقدر مقدار R به عدد یک نزدیک‌تر باشد، برآورد و برازش بهتر را نشان می‌دهد. ضریب کارایی

شرایط ایده آل برای داده‌های هیدرولوژیکی به ندرت ممکن است اتفاق بیفتد. در این مطالعه سطح‌های معنی‌داری  $P=0.05$  بکار گرفته شده است.

### نتایج و بحث

در این مطالعه، همانطور که اشاره شد داده‌های ۲۰ ساله (۱۳۹۷-۱۳۷۷) بارش و دبی ماهیانه از ۶ ایستگاه حوضه آجی‌چای دریافت گردید و بعد از متوسط‌گیری داده‌های بارش از ایستگاه‌های نامبرده (ارزنق، مهربان، مرکید، خواجه، ونیار و آخولا)، بارش ماهیانه بر روی حوضه محاسبه گردید و دبی خروجی از آخرین ایستگاه حوضه به نام آخولا جمع‌آوری شده و در نهایت از مدل‌های جنگل تصادفی (RF) و شبکه عصبی مصنوعی (MLP 2-5-1) بر پایه روش‌های داده‌کاوی به منظور شبیه‌سازی بارش-دبی ماهانه استفاده گردید. از مجموعه داده‌ها به صورت تصادفی ۷۰ درصد به عنوان داده‌های آزمون و ۳۰ درصد به عنوان داده‌های اعتبارسنجی استفاده شدند. سپس مجموعه آموزشی بارش به عنوان متغیر پیش‌بینی کننده و مجموعه آموزشی رواناب به عنوان متغیر هدف انتخاب شد. چندین ترکیب ورودی رواناب و بارندگی برای مدل‌سازی مورد ارزیابی قرار گرفت. داده‌های ورودی، مقادیر  $P$  و  $Q$  ماهانه قبلاً ثبت شده ( $P_t$ ،  $Q_{t-1}$ ) بوده و داده‌های خروجی مربوط به داده‌های رواناب فعلی ( $Q_t$ ) است (زیرنویس  $t$  نشان‌دهنده مرحله زمانی است). در نتیجه، چندین ترکیب ورودی از داده‌های  $P$  و  $Q$  ساخته شد (همانطور که در جدول ۲ ذکر شده است) و مدل RF و ANN (MLP 2-5-1) برای مدل‌سازی بارش-رواناب برای تعیین ترکیب ورودی بهینه استفاده شد.

این روش بطور متداول و گسترده‌ای در تحلیل روند سری‌های زمانی هیدرولوژیکی و هواشناسی بکار گرفته می‌شود (Lettenmaier 1994). از نقاط قوت این روش می‌توان به مناسب بودن کاربرد آن برای سری‌های زمانی‌ای که از توزیع آماری خاصی پیروی نمی‌کنند، اشاره کرد. اثرپذیری ناچیز این روش از مقادیر حدی که در برخی از سری‌های زمانی مشاهده می‌گردند نیز از دیگر مزیت‌های استفاده از این روش است (Turgay and Ercan 2005). فرض صفر این آزمون بر تصادفی بودن و وجود نداشتن روند در سری داده‌ها دلالت دارد و پذیرش فرض یک (رد فرض صفر) دال بر وجود روند در سری داده‌ها می‌باشد. آزمون من - کندال با تعریف آماره  $S$  به شکل زیر تعریف می‌شود:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(X_i - X_j) \quad (5)$$

که در آن  $X_i$  و  $X_j$  مقادیر مرتب شده‌ی نمونه و  $n$  تعداد نمونه است. مقدار  $\text{Sign}(x_j - x_i)$  برای  $(x_j - x_i) < 0$  برابر با منفی یک،  $(x_j - x_i) = 0$  برابر با صفر و برای  $(x_j - x_i) \geq 0$  برابر با یک است. مقدار نمره استاندارد شده  $Z$  و واریانس آماره  $S$  نیز از رابطه‌ی زیر محاسبه می‌شود:

$$Z_c = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & , S > 0 \\ 0 & , S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & , S < 0 \end{cases} \quad (6)$$

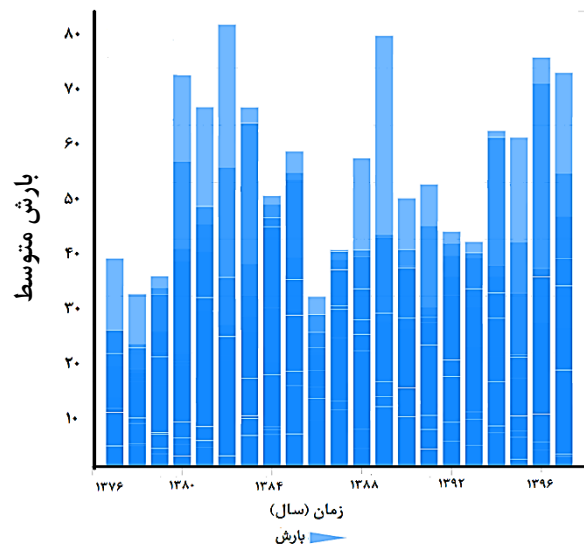
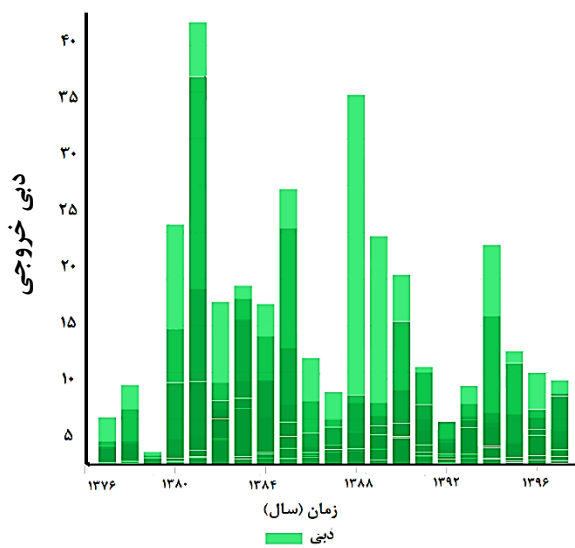
مقادیر مثبت  $Z$  نشان‌دهنده‌ی روندهای افزایشی و مقادیر منفی  $Z$  نشان‌دهنده‌ی روندهای کاهشی است. در شرایطی که  $S=0$  باشد داده‌ها فاقد روند بوده و هیچ روند معنی‌دار یا غیر معنی‌داری در داده‌ها وجود ندارد که این

جدول (۲): ترکیب ورودی از داده‌های  $P$  و  $Q$

ردیف	مدل	ورودی	خروجی
۱	جنگل تصادفی	$P_t, Q_{t-1}$	$Q_t$
۲	شبکه عصبی مصنوعی	$P_t, Q_{t-1}$	$Q_t$

دبی رودخانه آجی‌چای می‌باشد، مشاهده می‌شود مقدار آماره‌ی آزمون من‌کندال برای متغیر دبی در دوره زمانی ۱۳۹۴-۱۳۶۴ بترتیب میزان ۰/۳۶۹- بوده و در سطح ۹۵ درصد معنادار می‌باشد و نشان می‌دهد در دوره زمانی مورد مطالعه، روند کاهشی دبی معنادار بوده است. دبی خروجی در ایستگاه آخولا روندی نزولی دارد و در سال-های اخیر کاهش قابل ملاحظه‌ای در روند جریان خروجی از ایستگاه مورد مطالعه دیده می‌شود که زنگ خطری برای دریاچه ارومیه محسوب می‌شود. عوامل بسیاری در کاهش سطح آب دریاچه ارومیه دخیل می‌باشد که یکی از دلایل اصلی آن استفاده از آب رودخانه در کشاورزی و صنعت می‌باشد. دلیل بعدی را می‌توان احداث سد بر روی رودخانه‌های حوضه دریاچه ارومیه نام برد.

برای مشاهده سری زمانی بارش (الف) و رواناب (ب) ایستگاه آخولا در طی ۲۰ سال دوره مطالعاتی، نمودارهای (۴) در ادامه ارائه شده است. نمودار رسم شده در شکل (۴-الف) و مقدار آماره من‌کندال ۰/۳۴ که در سطح ۹۵ درصد معنی-دار نیست، مؤید این است که بارش سالانه در این بازه زمانی از نظر آماری معنادار نیست. بنابراین این گونه استیباط می‌شود که بارش سالانه در ناحیه مرکزی در سال‌های ۱۳۷۷ تا ۱۳۹۷ دارای روند نبوده و تغییرهای رخ داده در این دوره زمانی طبیعی می‌باشد و سری زمانی بارش نوسانی می‌باشد. (Salehi Babil et al. 2017) در مطالعات خود نشان دادند که نوسانات بارش در زیرحوضه‌های دریاچه ارومیه وجود دارد و روند مشخص افزایشی یا کاهشی در طی سال‌های اخیر در میزان بارندگی باران و برف بر روی حوضه مشاهده نشده است. اما در نمودار (۴-ب) که نشان دهنده



شکل (۴): سری زمانی بارش متوسط (الف) و دبی خروجی (ب) از ایستگاه آخولا

ستون سوم (تمام داده‌ها) شامل بخش‌های P و Q بود. در ستون چهارم و پنجم، مجموعه آموزشی و تست نشان داده شده است. بنابراین، مدل‌های RF و ANN(MLP 2-5-1) با داده‌های واقعی آموزش و آزمایش شدند. طبق جداول، میانگین بارش و دبی خروجی از ایستگاه برابر با ۲۱/۹۷ و ۱۱/۷۴ می‌باشد.

آمار بارندگی و رواناب (میانگین، حداقل، حداکثر) ایستگاه‌های آخولا در دوره مطالعاتی (۱۳۷۷-۱۳۹۷) در جدول (۳) ارائه شده است. همانطور که مشاهده می‌شود، ابتدا جداول به پنج ستون تقسیم شده و سپس ستون‌های سوم، چهارم و پنجم به دو قسمت تقسیم شدند.

جدول (۳): پارامترهای آماری داده‌های بارندگی و رواناب (۱۳۷۷-۱۳۹۷)

ایستگاه	پارامتر	داده‌های ورودی (۱۳۷۴-۱۳۵۵)		مجموعه آموزشی		مجموعه تست	
		Q	P	ANN	RF	ANN	RF
میانگین		۱۱/۷۴	۲۱/۹۷	۴/۰۳	۷/۶۲	۵/۳۳	۸/۱۵
حدافل		۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰
حداکثر		۹۲/۵۰	۷۴/۳۱	۳۹/۷۶	۶۷/۳۵	۴۷/۴۷	۶۷/۳۵

بالاترین کارایی و عملکرد را ۵ مین شبکه به نام MLP-2 5-1 ارائه داد.

به منظور نشان دادن قابلیت مدل‌ها برای مدل‌سازی بارش-رواناب، در این بخش فرآیند اعتبارسنجی مدل بهینه RF و ANN مورد بحث قرار می‌گیرد. تحلیل آماری همچنین برای تعیین همبستگی متقاطع بارندگی-رواناب با رواناب‌های پیشین به منظور نشان دادن عملکرد خوب این مدل‌ها برای ایستگاه آخولا در حوضه آجی‌چای انجام شد. آماره‌های همبستگی متقابل (Cross Correlation)، ضریب تعیین ( $R^2$ )، کارایی نش ساتکلیف (NSE) و خطای جذر میانگین مربعات (RMSE) بر اساس مجموعه داده‌های آموزش و آزمایش محاسبه شد (جدول ۴). نتایج نشان می‌دهد که مدل‌ها عملکرد خوبی دارند و جواب مناسبی داده است. در نهایت می‌توان با مقایسه نتایج بدست آمده برای هر دو مدل، بهترین مدل برای پیش‌بینی رواناب برای این ایستگاه را معرفی کرد.

نتایج مدل‌سازی رواناب با استفاده از مدل جنگل

تصادفی (RF) و شبکه عصبی مصنوعی (ANN)

در این مطالعه، برای مدل‌سازی رواناب در حوضه آجی‌چای با استفاده از مدل جنگل تصادفی و شبکه عصبی مصنوعی پرسپترون چند لایه (MLP) داده‌های بارندگی ماهانه، جریان ماهانه رودخانه، دمای متوسط ماهانه و تبخیر-تعرق ماهانه برآورد شده مورد استفاده قرار گرفت. مقادیر پنج معیار ارزیابی برای بررسی کمی عملکرد مدل RF در ایستگاه هیدرومتری لندی نیز در جدول (۴) ارائه شده است. در شبکه عصبی مصنوعی انواع ترکیب‌های ورودی داده‌ها با الگوریتم یادگیری BFGS<sup>۱۶</sup> با توابع عضویت سیگموئید و تانژانت هیپربولیک در لایه مخفی و تابع عضویت خطی در لایه خروجی با تعداد نرون‌های لایه مخفی متغیر بین ۲ تا ۱۰ نرون آموزش داده شد. این مدل چندین شبکه ی آموزش دیده را در هر نوع ترکیب ورودی تولید نمود که

جدول (۴): عملکرد مدل جنگل تصادفی در مجموعه آموزشی و تست

ایستگاه	داده‌های ورودی (۱۳۷۷-۱۳۹۷)	Cross Correlation		$R^2$		NSE		RMSE(m <sup>3</sup> /s)	
		ANN	RF	ANN	RF	ANN	RF	ANN	RF
آموزشی		۰/۷۶	۰/۸۳	۰/۵۸	۰/۶۹	۰/۵۸	۰/۶۲	۲/۸۱	۴/۵۸
آخولا	تست	۰/۸۶	۰/۸۲	۰/۷۴	۰/۶۷	۰/۷۳	۰/۵۵	۴/۲۴	۴/۸۸

به ترتیب برابر است با ۰/۶۲ و ۰/۵۵ و آخرین شاخص که در این مطالعه مورد بررسی قرار گرفته است خطای جذر میانگین مربعات می‌باشد که مقدار آن برای مجموعه‌های آموزشی و تست به ترتیب برابر با ۴/۵۸ و ۴/۸۸ بدست آمده است. اما در مدل ANN نتایج بدست آمده برای شاخص CC برای مجموعه‌های آموزشی و تست برابر با ۰/۷۶ و ۰/۸۶، مقادیر  $R^2$  برای داده‌های آموزشی و تست برابر با ۰/۵۸ و

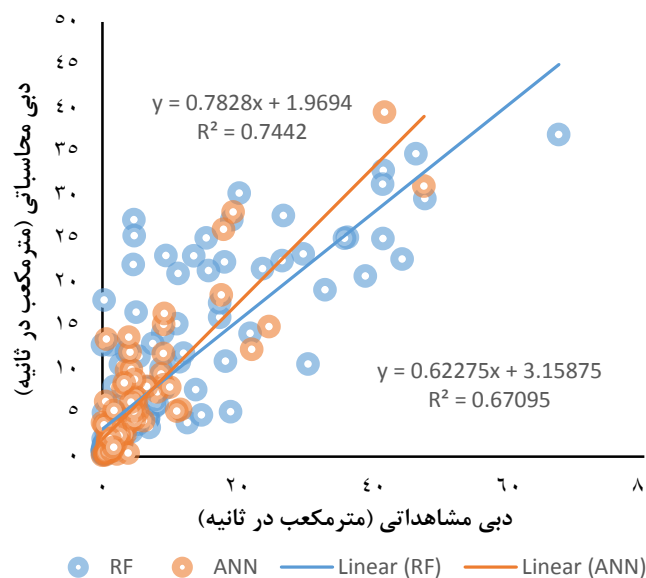
همانطور که در جدول بالا به وضوح دیده می‌شود در مدل جنگل تصادفی (RF)، همبستگی متقابل (Cross Correlation) در مجموعه آموزشی ۰/۸۳ و در مجموعه تست ۰/۸۲ بدست آمده است. عدد  $R^2$  در این مدل برای مجموعه‌های آموزشی و تست به ترتیب برابر با ۰/۶۹ و ۰/۶۷ تعیین شد. همچنین شاخص کارایی نش ساتکلیف بدست آمده برای ایستگاه آخولا برای مجموعه‌های آموزشی و تست

<sup>۱۶</sup>Broyden Fletcher Goldfarb Shanno

می‌شود با این که داده‌ها برای مجموعه آموزشی خوب آموزش دیده‌اند اما نتوانسته است برای مجموعه تست خوب جواب دهد، اما مدل ANN با اینکه برای مجموعه آموزشی عدد پایین‌تری نسبت به مدل RF دارد اما مجموعه تست این مدل دقت بالاتری نسبت به مدل جنگل تصادفی دارد و بنابراین برای مدل‌سازی بارش- رواناب مدل ANN مدل برتر مطالعه ما می‌باشد.

در شکل (۴) نمودار پراکنش مقادیر مشاهداتی در برابر مقادیر شبیه‌سازی شده دبی‌های ماهانه توسط مدل جنگل تصادفی و شبکه عصبی مصنوعی برای داده‌های بخش صحت سنجی نشان داده شد. با توجه به شکل‌های پراکنش پیش‌بینی دو مدل، مدل ANN نه تنها با درجه همبستگی (R) بیشتری فرآیند بارش- رواناب ماهانه را در این تحقیق شبیه‌سازی نموده است، بلکه تا حدودی در برآورد مقادیر حداکثر رواناب عملکرد بهتری داشته است.

شاخص NSE نیز برای مجموعه‌های آموزشی و تست به ترتیب برابر با ۰/۵۸ و ۰/۷۳ و آخرین شاخص RMSE نیز برای دو مجموعه‌های آموزشی و تست برابر با ۳/۸۱ و ۴/۲۴ تعیین گردید. حال با توجه به اینکه در تحلیل مقالات از داده‌های تست برای بررسی عملکرد مدل‌ها استفاده می‌شود، می‌توان گفت در این مطالعه نیز  $CC$ ،  $R^2$ ،  $NSE$  و  $RMSE$  بدست آمده در داده‌های تست، برای بارش و رواناب اعداد قابل قبولی هستند. با توجه به نتایج جدول (۴) مشخص است که در این دوره مطالعاتی برای ایستگاه آخولا مدل ANN نسبت به مدل RF عملکرد بهتری داشته است و با دقت خوبی توانسته دبی جریان و بارش ماهانه را شبیه‌سازی کند. همچنین در این دوره‌ها در ایستگاه پل سنخ مدل RF کارایی بهتری را نسبت به مدل ANN نشان می‌دهد. نکته‌ای که قابل ذکر است بیش برآزش (overfit) شدن مدل RF می‌باشد، همانطور که در جدول (۴) مشاهده



شکل (۴): نمودارهای پراکنده رواناب شبیه‌سازی شده با استفاده از RF و ANN ایستگاه آخولا در دوره مطالعاتی (۱۳۹۷-۱۳۹۷)

گردید. مقدار بارش و رواناب با گام تاخیر زمانی یک روزه بعنوان ورودی به این مدل‌ها وارد شد. سپس مقادیر رواناب ماهانه مشاهداتی با رواناب ماهانه تخمین زده شده توسط مدل‌ها با استفاده از معیارهای ارزیابی مورد بررسی گرفت. نتایج نشان داد که مدل‌ها از عملکرد قابل قبولی در برآورد رواناب ماهانه برخوردار بودند. در این مطالعه مقادیر  $CC$ ،  $R^2$ ،  $NSE$  و  $RMSE$  برای مجموعه تست در مدل

### نتیجه‌گیری

در این تحقیق کارایی مدل‌های هوشمند به منظور برآورد میزان رواناب ماهانه حوضه آبی‌چای مورد بررسی قرار گرفت. در این مطالعه جهت پیش‌بینی جریان ماهانه رودخانه آبی‌چای در ایستگاه هیدرومتری آخولا، از نرم افزار STATISTICA و مدل جنگل تصادفی (RF) و شبکه عصبی مصنوعی پرسپترون چندلایه (MLP 2-5-1) استفاده

زمانی دبی رودخانه آبی‌چای در ایستگاه آخولا نشان داد که روند کاملاً نزولی برای جریان آب رودخانه در این مدت ثبت شده است که در واقع دلیل کاهش دبی ورودی به دریاچه ارومیه می‌باشد. از دلایل مهم کاهش جریان ورودی به دریاچه و کاهش تراز آب در سال‌های اخیر، استفاده از جریان آب در بخش کشاورزی و صنعت می‌باشد. همین‌طور احداث تعداد بیشمار سد در حوضه آبی‌چای نیز می‌تواند جز دلایل مهم باشد. با بررسی نتایج این مطالعه می‌توان گفت با توجه به عملکرد قابل قبول مدل‌سازی بارش - رواناب برای حوضه آبی‌چای، برای بقیه ایستگاه‌های زیرحوضه دریاچه ارومیه نیز می‌توان از مدل شبکه عصبی مصنوعی استفاده کرد.

جنگل تصادفی به ترتیب برابر با ۰/۸۲، ۰/۶۷، ۰/۵۵ و ۴/۸۸ و در مدل ANN نتایج بدست آمده برای شاخص‌های  $CC$ ،  $R^2$ ، NSE و RMSE در مجموعه تست به ترتیب برابر با ۰/۸۶، ۰/۷۴، ۰/۷۳ و ۴/۲۴ تعیین گردید. تحلیل‌های نتایج نشان داد که با توجه به عملکرد بالای مجموعه تست در مدل شبکه عصبی مصنوعی در ایستگاه آخولا، این مدل کارایی بالاتری نسبت به مدل RF دارد. از نتایج دیگر این پژوهش می‌توان به سری زمانی بارش و رواناب طی ۲۰ سال اخیر اشاره کرد. در این مطالعه تحلیل سری‌های زمانی با استفاده از آزمون ناپارامتریک من-کندال انجام شد، نتایج نشان داد که در طی این مدت روند مشخصی برای بارش بر روی حوضه آبی‌چای دیده نشد. در واقع نمودار نشان داد که بارش در این مناطق بصورت نوسانی بوده است. اما سری

#### منابع

1. Adnan R.M., Yuan X., Kisi O., Adnan M., and Mehmood A. 2018. Stream flow forecasting of poorly gauged mountainous watershed by least square support vector machine, fuzzy genetic algorithm and M5 model tree using climatic data from nearby station. *Water Resources Management*, 32(14): 4469-4486. <https://doi.org/10.1016/j.jhydrol.2019.124371>.
2. Breiman L. 1984. Classification and regression trees CA, wadsworth International Groups. *Handbook of Data Mining and Knowledge Discovery*, 248-276.
3. Breiman L. 2001. Random forests. *Machine Learning*, 45(1):5-32. <http://dx.doi.org/10.1023/A:1010933404324>
4. Catani F., Lagomarsino D., Segoni S., and Tofani V. 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards Earth System Science*, 13:2815-2831. <https://doi.org/10.5194/nhess-13-2815-2013>
5. Dastorani M.T., Mahjoobi J., Talebi A., and Fakhar F. 2018. Application of machine learning approaches in rainfall-runoff modeling (Case study: Zayandeh rood basin in Iran). *Civil Engineering Infrastructures Journal*, 51(2):293-310. <http://dx.doi.org/10.7508/CEIJ.2018.02.004>
6. Dayhoff J.E. 1990. *Neural network principles*. Prentice-Hall International. U.S.A. 197. [https://doi.org/10.1002/1097-0142\(20010415\)91:8+<1615::AID-CNCR1175>3.0.CO;2-L](https://doi.org/10.1002/1097-0142(20010415)91:8+<1615::AID-CNCR1175>3.0.CO;2-L)
7. Ebrahimi Mohammadi S.H., Boshri M., and Ghaleh S. 2011. Modeling and prediction of monthly discharge stream (Case study: Qarasou river) 4<sup>th</sup> Iran Water resources management conference, Amir Kabir University of Technology Tehran (In Persian).
8. Firat M., and Gungor M. 2006. River flow estimation using adaptive neuro-fuzzy inference system. *Mathematics and Computers in Simulation*, 75:87-96. <https://doi.org/10.1016/j.matcom.2006.09.003>
9. Hassanzadeh Y., Abdi Kordani A., Shafiei Najd M., and Khoshtinat S. 2015. Daily streamflow forecasting of Nooranchay river using the hybrid model of artificial neural networks- principal component analysis. *Water and Soil Science*, 25(3):53-63 (In persian). <https://doi.org/10.1016/j.matcom.2006.09.003>

10. Hosseini H. 1999. Artificial neural network model for determining flood hydrograph. Master's Thesis in Civil Engineering. Iran University of Science and Technology. 103. <https://doi.org/10.1002/widm.8>
11. Hussain D., and Khan, A.A. 2020. Machine learning techniques for monthly river flow forecasting of Hunza river, Pakistan. *Earth Science Informatics*, 13: 939-949. <https://doi.org/10.1007/s12145-020-00450-z>
12. Kendall M.G. 1975. Rank correlation methods, fourth ed. Charles Griffin, London.
13. Khademi M., Fazloulou R., and Emadi A. 2017. Investigation the ability of artificial neural network in simulation of rainfall-runoff process under the climate change conditions (Case study: Pashakola Babol Dam Basin). *Journal of Watershed Management Research*, 8 (16) :53-64. <http://jwmr.sanru.ac.ir/article-1-903-en.html>
14. Khanna T. 1990. Foundation of neural networks. Addison-Wesley Publishing Company. U.S.A. 327.
15. Kisi O. 2010. Wavelet regression model for short-term streamflow forecasting, *Journal of Hydrology*, 389: 344-353. <https://doi.org/10.1016/j.jhydrol.2010.06.013>
16. Kurtulus B., and Razack M. 2010. Modeling daily discharge responses of a large karstic aquifer using soft computing methods: Artificial neural network and neuro fuzzy, *Journal of Hydrology*, 381: 101-111. <https://doi.org/10.1016/j.jhydrol.2009.11.029>
17. Lettenmaier D.P., Wood E.F. and Wallis J.R. 1994. Hydro-climatological trends in the continental United States, 1948–88. *Journal of Climate*, 7: 586–607. <https://www.jstor.org/stable/26197880>
18. Lorrai M., and Sechi M.G. 1995. Neural nets for modeling rainfall-runoff transformation, *Water Resources Management*, 9:299-313. <https://doi.org/10.1007/BF00872489>
19. Luce C. 2014. Runoff prediction in ungauged basins: synthesis across processes, places and scales: edited by günter blöschl, murugesu sivapalan, Thorsten Wagener, Alberto Viglione, and Hubert Savenije Cambridge University Press, 2013, 465 pp., ISBN: 978-1107028180, 140(hardback), 112 (eBook). <https://doi.org/10.1002/2014EO020025>
20. Mann H.B. 1945. Non-parametric test against trend. *Journal of Econometrical*, 13, 245-259.
21. Nayak P.C., Sudheer K.P., Rangan D.M. and Ramasastri K.S. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1-2):52-66. <https://doi.org/10.1016/j.jhydrol.2003.12.010>
22. Nicodemus K.K. 2011. Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*. 12: 369-373.
23. Nilsson P., Uvo C.B., and Bentsen R. 2005. Monthly runoff simulation: comparing and combining conceptual and neural network models, *Journal of Hydrology*, 321: 344-363. <https://doi.org/10.1016/j.jhydrol.2005.08.007>
24. Nourani V., M Keynezhad., and Makani L. 2009. Using adaptive neuro-fuzzy inference system rainfall-runoff modeling. *Journal of Civil and Environmental Engineering*, 39: 75-81 (In Persian). <https://civilica.com/doc/63080>
25. Pham L.T., Luo L., and Finley A. 2021. Evaluation of random forests for short-term daily streamflow forecasting in rainfall-and snowmelt-driven watersheds. *Hydrology and Earth System Sciences*, 25(6):2997-3015. <https://doi.org/10.5194/hess-25-2997-2021>

26. Phomcha P., Wirojanagud P., Vangpaisal T., and Thaveevouthti T. 2011. Suitability of SWAT model for simulating of monthly streamflow in lam sonthi watershed. *The Journal of Industrial Technology*, 7(2): 49- 56.
27. Poursalehi F., KhasheiSiuki A., and Hashemi S.R. 2022. Investigating the performance of random forest algorithm in predicting water table fluctuations compared with two models of decision tree and artificial neural network (Case study: unconfined aquifer of Birjand plain). *Iranian Journal of Ecohydrology*, 8(4): 961-974 (In Persian). <https://doi.org/10.5194/hess-25-2997-2021>
28. Razavi T., and Coulibaly P. 2013. Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering*, 18(8): 958-975. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000690](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690)
29. Sabouhi, R. and Soltani, S., 2008. Analysis of the climate trend in the major cities of Iran. *Journal of Agricultural Science and Technology*. 12(46):303-321 (In Persian).
30. Salehi Babil S, Zeinalzadeh K., and Hessari B. 2017. The changes in the frequency of daily precipitation in Urmia Lake basin, Iran. *Journal of Theoretical and Applied Climatology*, 1- 10. <https://doi.org/10.1007/s00704-017-2177-7>
31. Serrano A., Mateos V.L., and Garcia J.A. 1999. Trend analysis of monthly precipitation over the Iberian Peninsula for the period 1921-1995. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*. 24(1-2): 85-90. [https://doi.org/10.1016/S1464-1909\(98\)00016-1](https://doi.org/10.1016/S1464-1909(98)00016-1)
32. Shafeizadeh M., Fathian H., and Nikbakht Shahbazi A. 2019. Continuous rainfall-runoff simulation by artificial neural networks based on efficient input variables selection using partial mutual information (PMI) algorithm. *Iran Water Resources Research*, 15(2): 144-161. (In Persian). <https://doi.org/20.1001.1.17352347.1398.15.2.12.1>
33. Sohrabi Geshnigani F., Mirabbasi Najafabadi R., and Golabi M.R. 2021. Rainfall-Runoff modeling using HBV model and random forest algorithm in Bazfat watershed. *Iranian Journal of Soil and Water Research*, 52(5):1395-1407.( In Persian). <https://doi.org/10.22059/IJSWR.2021.318980.668893>
34. Toker A.S., and Markus M. 2000. Precipitation-runoff modeling using artificial neural network and conceptual models, *Journal of Hydrologic Engineering*, 5: 156-161. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(156\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(156))
35. Turgay P., and Ercan K. 2005. Trend analysis in turkish precipitation data. *Hydrological Processes: An International Journal*. 20(9): 2011-2026. <https://doi.org/10.1002/hyp.5993>
36. Worland S.C., Farmer W.H., and Kiang J.E. 2018. Improving predictions of hydrological low-flow indices in ungagged basins using machine learning. *Environmental Modelling and Software*, 101:169-182. <https://doi.org/10.1016/j.envsoft.2017.12.021>
37. Youssef A.M., Pourghasemi H.R., Pourtaghi Z.S., and Al-Katheeri M.M. 2015. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region. *Saudi Arabia Landslides*. <https://doi.org/10.1007/s10346-015-0614-1>.
38. Zarei M., Zandi R., and Naemitabar M. 2022. Assessment of flood occurrence potential using data mining models of support vector machine, chaid and random forest (Case

study: Frizi watershed). Journal of Wound Management and Research, 13 (25) :133-144. (In Persian). <http://jwmr.sanru.ac.ir/article-1-1140-en.html>.

## **Rainfall-Runoff Modeling of Aji Chai Basin Using Random Forest and Artificial Neural Network Models**

**Zeinab Bigdeli\*<sup>1</sup>, Abolfazl Majnooni-Heris<sup>2</sup>, Reza Delirhasannia<sup>2</sup>, Sepideh Karimi<sup>3</sup>**

DOI: 10.22103/nrswe.2023.20278.1013

### **Abstract**

Estimating runoff due to rainfall is a very important step in planning water resources, especially in watersheds without hydrometric stations. In this study, the rainfall-runoff simulation of Akhola station located in the Ajichai basin was discussed and the most suitable rainfall-runoff model was presented by using data mining methods and comparing their performance. For this purpose, the desired data (rainfall, discharge, temperature) for this study were received monthly from the water and meteorological organizations of the East and West Azerbaijan provinces. Random Forest and Artificial Neural Network data mining models were used for simulation. The comparison of observed monthly runoff values with monthly runoff estimated by models was done using evaluation criteria. In this study, CC values (correlation coefficient) for test sets in the random forest model and artificial neural network were determined as 0.77 and 0.86, respectively. The analysis of the results showed that the ANN model has a higher performance and efficiency than the RF model for the Akhola station. Among other results of this research, we can mention the time series of rainfall and runoff of the station during the last 20 years. According to the obtained graphs, during these 20 years, there was no clear trend for precipitation in the Ajichai basin, and the time series graphs showed that the precipitation in these areas was fluctuating. But the time series for Ajichai discharge at Akhula station showed that a completely downward trend was recorded for the river water flow, which is the reason for the decrease of the discharge entering Lake Urmia and the lowering of the lake water level.

**Keywords:** Rainfall-Runoff, Random Forest, Aji Chai Basin, Modeling, Machine Learning

<sup>1\*</sup>- Ph.D Candidate, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.  
Corresponding author, email: Z.bigdeli.20@gmail.com

<sup>2</sup> - Associate Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.

<sup>3</sup> - Post Ph.D Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.