

مدل سازی بارش ماهانه با استفاده از مدل های هوشمند یادگیری ماشین و یادگیری عمیق (مطالعه موردی: دشت مغان)

پویا اللهویردی پور^{۱*}، یعقوب دین پژوه^۲

۱- دانشجوی دکتری مهندسی منابع آب، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

a.pouya1996@gmail.com

۲- استاد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. dinpashoh@yahoo.com

چکیده

بارش به عنوان یکی از مهم ترین مؤلفه های چرخه هیدرولوژیکی، نقش کلیدی در مدیریت منابع آب و کشاورزی در مناطق حساس مانند دشت مغان ایفا می کند. این مطالعه با هدف مدل سازی بارش های ماهانه در دشت مغان با استفاده از الگوریتم های یادگیری ماشین و یادگیری عمیق شامل شبکه عصبی طولانی کوتاه مدت (LSTM)، درختان اضافی (ET)، ماشین بردار پشتیبان (SVM) و شبکه عصبی پرسپترون چندلایه (MLP) انجام شد. داده های هواشناسی شامل بارش، دمای متوسط، رطوبت متوسط و سرعت باد متوسط بلندمدت شامل ۳۶۰ رکورد ماهانه از سال ۱۹۹۳ تا ۲۰۲۲ گردآوری شده و به دو بخش تقسیم شدند: ۲۵۲ رکورد برای آموزش (۷۰ درصد) و ۱۰۸ رکورد برای تست (۳۰ درصد). عملکرد مدل ها با معیارهای ضریب همبستگی (r)، جذر میانگین مربعات خطاها (RMSE) و ضریب نش-ساتکلیف (NSE) ارزیابی شد. در بخش آموزش، مدل LSTM با $r = 0.91$ و $NSE = 0.87$ ، $RMSE = 8.23$ mm بهترین عملکرد را نشان داد. در بخش صحت سنجی نیز مدل LSTM با $r = 0.89$ و $NSE = 0.84$ ، $RMSE = 8.41$ mm بهترین عملکرد را داشت، در حالی که مدل ET با $r = 0.86$ و $NSE = 0.79$ ، $RMSE = 8.79$ mm و مدل SVM با $r = 0.81$ و $NSE = 0.62$ ، $RMSE = 10.23$ mm و مدل MLP با $r = 0.83$ و $NSE = 0.70$ ، $RMSE = 9.48$ mm ضعیف تری نشان دادند. تحلیل نمودارهای پراکنش نشان داد که همه مدل ها در پیش بینی بارش های نزدیک به صفر با چالش مواجه هستند. نتایج حاکی از برتری LSTM در هر دو بخش آموزش و تست است، که نشان دهنده پتانسیل بالای یادگیری عمیق در پیش بینی بارش و حمایت از مدیریت پایدار منابع آب است.

واژه های کلیدی: بردار پشتیبان، درختان اضافی، شبکه عصبی، منابع آب، هوش مصنوعی، یادگیری ماشین.

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

۱- مقدمه

بارش به عنوان یکی از مهم‌ترین مؤلفه‌های چرخه هیدرولوژیکی، نقش کلیدی در تأمین منابع آب، تولیدات کشاورزی و پایداری اکوسیستم‌ها ایفا می‌کند. تغییرات در الگوهای بارش، به‌ویژه در مناطق حساس به تغییرات اقلیمی، می‌تواند تأثیرات عمیقی بر مدیریت منابع آب و برنامه‌ریزی کشاورزی داشته باشد. در مناطقی مانند دشت مغان، که به دلیل موقعیت جغرافیایی و وابستگی زیاد به منابع آبی برای کشاورزی، از اهمیت ویژه‌ای برخوردار است، پیش‌بینی دقیق بارش‌های ماهانه برای کاهش ریسک‌های مرتبط با کمبود آب و بهینه‌سازی مدیریت منابع ضروری است [1]. تغییرات اقلیمی، با ایجاد نوسانات در الگوهای بارش و افزایش دفعات رویدادهای شدید و فرین‌های هواشناسی، چالش‌های جدیدی را در مدل‌سازی و پیش‌بینی بارش ایجاد کرده است [2, 3].

مدل‌سازی بارش‌های ماهانه به دلیل ماهیت غیرخطی و پیچیده داده‌های هواشناسی، همواره یکی از موضوعات چالش‌برانگیز در علوم هیدرولوژی و اقلیم‌شناسی بوده است. روش‌های سنتی مانند مدل‌های رگرسیونی و آماری، به دلیل محدودیت در ثبت دینامیک‌های غیرخطی و وابستگی‌های پیچیده زمانی و مکانی، اغلب دقت محدودی در پیش‌بینی بارش دارند. این محدودیت‌ها، به‌ویژه در مناطقی مانند دشت مغان که تحت تأثیر عوامل متعددی از جمله تغییرات فصلی، توپوگرافی و فعالیت‌های انسانی قرار دارد، به کاهش کارایی این مدل‌ها منجر می‌شود [4].

در سال‌های اخیر، پیشرفت‌های قابل‌توجه در حوزه یادگیری ماشین، امکان مدل‌سازی دقیق‌تر و انعطاف‌پذیرتر داده‌های هواشناسی را فراهم کرده است. الگوریتم‌های یادگیری ماشین و یادگیری عمیق، مانند شبکه‌های عصبی مصنوعی (ANN)، ماشین‌های بردار پشتیبان (SVM)، جنگل تصادفی (RF)، حافظه بلند کوتاه‌مدت (LSTM) و مدل‌های مبتنی بر درخت تصمیم، به دلیل توانایی در شناسایی الگوهای پیچیده و غیرخطی در داده‌ها، به‌طور گسترده‌ای در پیش‌بینی متغیرهای هواشناسی و هیدرولوژیکی مورد استفاده قرار گرفته‌اند. این روش‌ها می‌توانند با بهره‌گیری از داده‌های بلندمدت هواشناسی، دقت پیش‌بینی بارش را بهبود بخشیده و ابزارهای قدرتمندی برای مدیریت منابع آب ارائه دهند [5-11].

این مطالعه با هدف مدل‌سازی بارش‌های ماهانه در دشت مغان با استفاده از الگوریتم‌های پیشرفته و هوشمند یادگیری ماشین و یادگیری عمیق شامل حافظه بلند کوتاه‌مدت (LSTM)، ماشین بردار پشتیبان (SVM)، شبکه عصبی پرسپترون چندلایه (MLP) و مدل درختان اضافی (ET) انجام شده است. با استفاده از داده‌های هواشناسی بلندمدت و ارزیابی عملکرد مدل‌های مختلف، این پژوهش به دنبال ارائه رویکردی علمی برای پیش‌بینی دقیق بارش‌های ماهانه و حمایت از مدیریت پایدار منابع آب در این منطقه استراتژیک است.

۲- مواد و روش‌ها

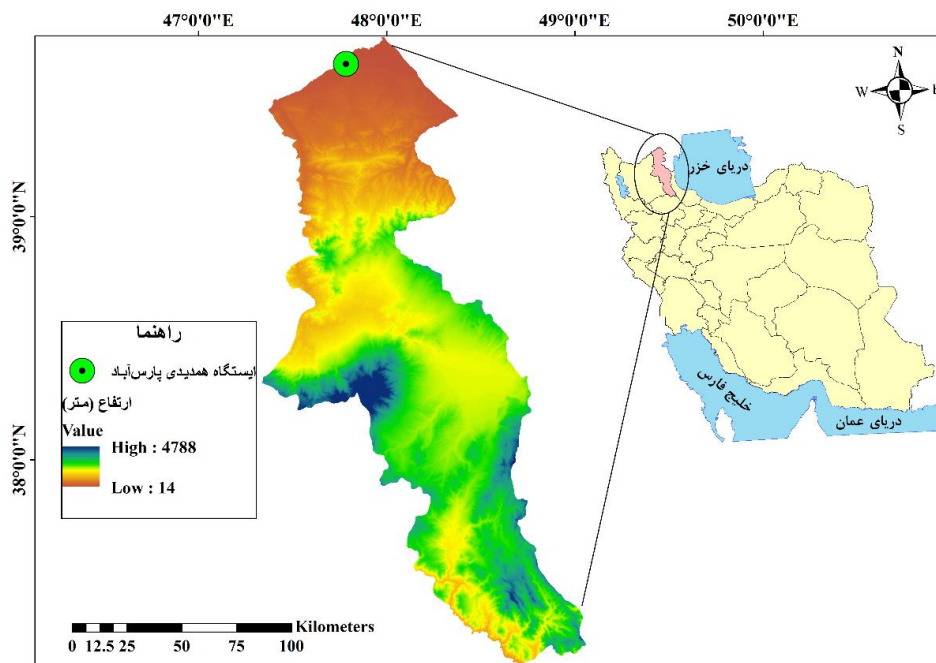
۲-۱- منطقه مورد مطالعه و داده‌های مورد استفاده

دشت مغان، واقع در بخش شمالی استان اردبیل و در مجاورت مرز ایران و جمهوری آذربایجان، یکی از مهم‌ترین مناطق کشاورزی ایران است. این دشت با مساحتی بالغ بر ۲۰۰,۰۰۰ هکتار، نقش کلیدی در تولید محصولات کشاورزی و تأمین امنیت غذایی کشور ایفا می‌کند. خاک‌های حاصلخیز، دسترسی به منابع آبی از جمله رودخانه ارس، و شرایط اقلیمی مناسب، این منطقه را به یکی از مراکز اصلی کشاورزی در شمال غرب ایران تبدیل کرده است. با این حال، تغییرات اقلیمی و الگوهای نامنظم بارش در سال‌های اخیر، چالش‌هایی را برای پایداری تولید کشاورزی در این منطقه ایجاد کرده است. در این مطالعه، داده‌های هواشناسی از ایستگاه همدید پارس‌آباد مغان، به‌عنوان نماینده شرایط اقلیمی منطقه، مورد استفاده قرار گرفته است. این ایستگاه

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

در مختصات جغرافیایی تقریبی ۳۹ درجه و ۳۶ دقیقه شمالی و ۴۷ درجه و ۴۷ دقیقه شرقی و در ارتفاع ۷۲ متری از سطح دریا واقع شده است (شکل ۱).

بر اساس داده‌های بلندمدت، میانگین بارش سالانه در این ایستگاه حدود ۲۶۱ میلی‌متر و میانگین دمای سالانه ۱۵/۵ درجه سانتی‌گراد است. این مشخصات اقلیمی، همراه با تنوع فصلی در بارش و دما، دشت مغان را به یک منطقه ایده‌آل برای مطالعه الگوهای اقلیمی تبدیل کرده است. داده‌های ایستگاه پارس‌آباد به دلیل پوشش طولانی‌مدت و دقت بالا، مبنای مناسبی برای تحلیل‌های هیدرولوژیکی و پیش‌بینی‌ها فراهم می‌کنند.



شکل ۱: موقعیت جغرافیایی منطقه مورد مطالعه، دشت مغان، استان اردبیل، ایران

در این پژوهش، داده‌های مورد استفاده برای مدل‌سازی بارش ماهانه در دشت مغان، از ایستگاه همدید پارس‌آباد مغان در یک بازه زمانی ۳۰ ساله (۱۹۹۳-۲۰۲۲) گردآوری شده‌اند. این دوره طولانی‌مدت، امکان تحلیل دقیق الگوهای بارش را فراهم می‌کند. متغیرهای اقلیمی بارش در یک ماه قبل، بارش در دو ماه قبل، دمای یک ماه قبل، رطوبت نسبی یک ماه قبل و سرعت باد یک ماه قبل به‌عنوان متغیرهای ورودی مدل‌های هوشمند برای پیش‌بینی بارش ماه آتی انتخاب شده‌اند، زیرا این متغیرها به‌طور مستقیم بر الگوهای بارش تأثیر می‌گذارند [7].

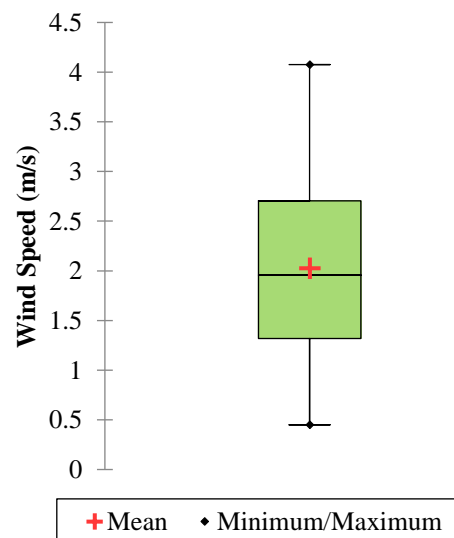
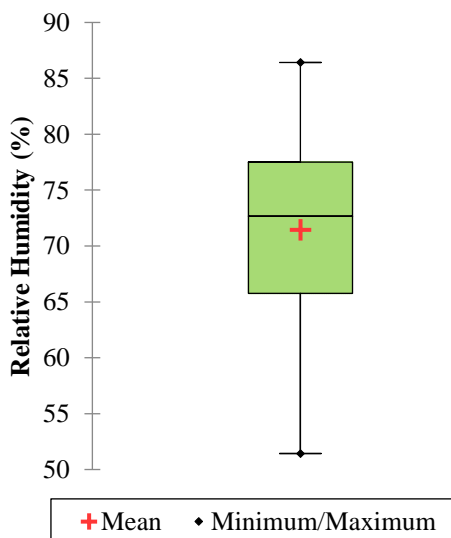
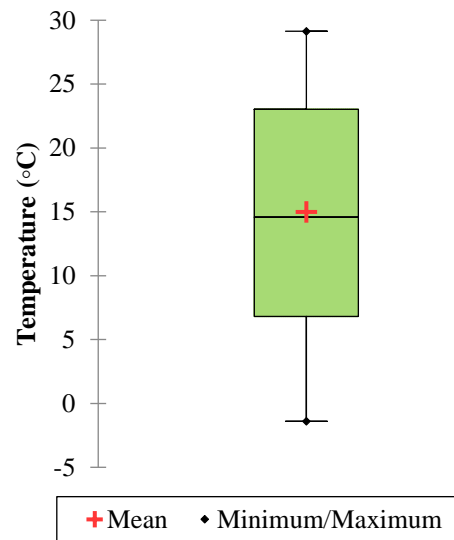
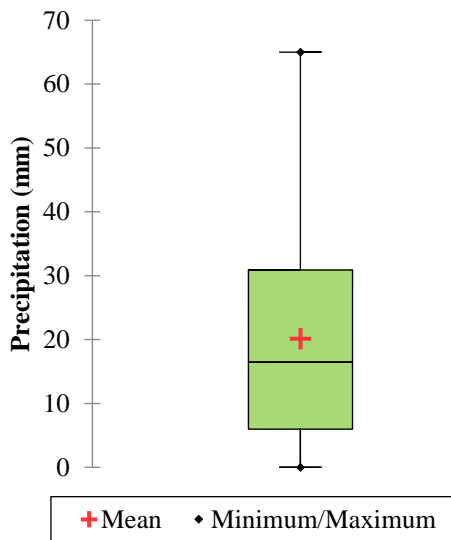
مشخصات آماری متغیرهای مورد استفاده شامل بارش (میلی‌متر)، دما (درجه سانتی‌گراد)، رطوبت نسبی (درصد)، و سرعت باد (متر بر ثانیه) در جدول ۱ ارائه شده است، که مبنای تحلیل‌های بعدی در این مطالعه قرار گرفته است. در شکل ۲ نیز نمودارهای جعبه‌ای متغیرهای مورد استفاده آمده است. این مشخصات آماری نشان‌دهنده تنوع قابل توجه در متغیرهای اقلیمی دشت مغان است که پیچیدگی‌های شبیه‌سازی بارش ماهانه را برجسته می‌کند. انحراف معیار بالای بارش و دما نشان‌دهنده تغییرات اقلیمی قابل توجه در منطقه است که نیاز به مدل‌های هوشمند و پیشرفته یادگیری ماشین مانند SVM، MLP، LSTM و ET را برای ثبت روابط غیرخطی بین متغیرهای ورودی و بارش ماه آتی تأیید می‌کند. این تحلیل اولیه، مبنای

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

مناسبی برای ارزیابی عملکرد مدل‌های یادگیری ماشین در پیش‌بینی بارش فراهم می‌کند و بر اهمیت انتخاب متغیرهای ورودی مرتبط با الگوهای اقلیمی تأکید دارد.

جدول ۱: مشخصات آماری متغیرهای اقلیمی دشت مغان در مقیاس ماهانه در دوره ۱۹۹۳-۲۰۲۲

متغیر	واحد	کمینه	بیشینه	میانگین	انحراف معیار
بارش	میلی‌متر	۰	۶۵	۲۰/۱۴	۱۶/۷۴
دمای متوسط	درجه سانتی‌گراد	-۱/۴۰	۲۹/۱۴	۱۴/۹۹	۸/۶۳
رطوبت نسبی متوسط	درصد	۵۱/۴۳	۸۶/۴۲	۷۱/۴۶	۷/۷۸
سرعت باد متوسط	متر بر ثانیه	۰/۴۵	۴/۰۸	۲/۰۳	۰/۸۶



شکل ۲: نمودار جعبه‌ای متغیرهای اقلیمی دشت مغان، مورد استفاده در تحقیق حاضر (دوره ۱۹۹۳-۲۰۲۲)

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

۲-۲- ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM) یک الگوریتم یادگیری ماشین نظارت شده است که به طور گسترده در مسائل طبقه بندی و رگرسیون، از جمله مدل سازی اقلیمی و هیدرولوژیکی استفاده می شود. این روش با یافتن یک ابرصفحه بهینه که داده ها را با حداکثر حاشیه (margin) جدا می کند، دقت پیش بینی را بهبود می بخشد و در برابر بیش برآزش (overfitting) مقاوم است. SVM بر پایه مفهوم حداکثر کردن حاشیه (margin maximization) عمل می کند، به این معنا که ابرصفحه ای (hyperplane) را پیدا می کند که داده های دو کلاس را با بیشترین فاصله ممکن از نقاط مرزی (support vectors) جدا کند [12].

در مسائل رگرسیون (SVR یا Support Vector Regression)، هدف یافتن یک تابع است که مقادیر را با حداقل خطا و در عین حال حفظ یک حاشیه مشخص پیش بینی کند. در فضای ویژگی ها، SVM یک ابرصفحه تعریف می کند که داده ها را به دو دسته تقسیم می کند. برای داده های غیرخطی، از ترفند کرنل (kernel trick) استفاده می شود تا داده ها به فضای با ابعاد بالاتر نگاشت شوند، جایی که یک ابرصفحه خطی می تواند داده ها را جدا کند. کرنل هایی مانند خطی، چندجمله ای، یا گاوسی (RBF - Radial Basis Function) برای مدیریت روابط غیرخطی بین متغیرها استفاده می شوند. کرنل RBF به دلیل انعطاف پذیری در مدل سازی داده های پیچیده، پر کاربرد است. پارامترهایی مانند C (جریمه خطا) و γ (پارامتر کرنل RBF) برای کنترل تعادل بین پیچیدگی مدل و خطای آموزش استفاده می شوند. SVM با استفاده از داده های آموزشی، ابرصفحه بهینه را محاسبه می کند که حاشیه را حداکثر کرده و خطاها را به حداقل می رساند. در نهایت، مدل آموزش دیده برای پیش بینی مقادیر جدید در مجموعه آزمایشی استفاده می شود [13].

۲-۳- درختان اضافی (ET)

مدل درختان اضافی (Extra Trees یا Extremely Randomized Trees) یک روش یادگیری جمعی (Ensemble Learning) مبتنی بر درخت های تصمیم گیری است که به عنوان یک توسعه از الگوریتم جنگل تصادفی (Random Forest) معرفی شده است. این روش با هدف بهبود دقت پیش بینی و کاهش پیچیدگی محاسباتی طراحی شده است و به ویژه در مسائل رگرسیون و طبقه بندی، از جمله مدل سازی هیدرولوژیکی، کاربرد گسترده ای دارد. مدل Extra Trees مشابه جنگل تصادفی از مجموعه ای از درخت های تصمیم گیری تشکیل شده است که به صورت موازی آموزش داده می شوند. با این حال، دو تفاوت کلیدی در مکانیزم آن وجود دارد که آن را از جنگل تصادفی متمایز می کند:

۱- در جنگل تصادفی، برای هر گره در درخت تصمیم گیری، بهترین نقطه تقسیم از میان زیرمجموعه ای تصادفی از ویژگی ها انتخاب می شود. اما در Extra Trees، نقاط تقسیم (split points) به صورت کاملاً تصادفی از میان ویژگی ها و مقادیر ممکن انتخاب می شوند، بدون بررسی بهینه بودن تقسیم. این رویکرد تصادفی سازی قوی تر، واریانس مدل را کاهش می دهد و آن را در برابر بیش برآزش (overfitting) مقاوم تر می کند.

۲- استفاده از کل داده های آموزشی: برخلاف جنگل تصادفی که از روش نمونه گیری بوت استرپ (Bootstrap Sampling) برای ایجاد زیرمجموعه های داده استفاده می کند، Extra Trees معمولاً از کل مجموعه داده های آموزشی برای ساخت هر درخت استفاده می کند، که این امر زمان محاسبات را کاهش می دهد.

از مزایای مدل ET میتوان به کارایی محاسباتی (به دلیل حذف مرحله جستجوی بهینه برای نقاط تقسیم)، مقاومت در برابر بیش برآزش (به دلیل تصادفی سازی قوی تر)، انعطاف پذیری (مخصوصاً برای داده های با ابعاد بالا و متغیرهای چندگانه مانند متغیرهای هواشناسی) و عملکرد قوی در بسیاری از مسائل (به ویژه در داده های نویزی) اشاره کرد [14].

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

۲-۴- مدل حافظه طولانی کوتاه مدت LSTM

یکی از انواع شبکه‌های عصبی بازگشتی روش LSTM می‌باشد که برای پیش‌بینی سری‌زمانی با یادگیری طولانی مدت استفاده می‌شوند. این یک روش خاص عصبی بازگشتی می‌باشد که برای مدل‌سازی توالی و وابستگی‌های طولانی مدت ساخته شده‌است. LSTM یک تکامل با کیفیت بالا از شبکه‌های عصبی بازگشتی (RNN) می‌باشد و به دلیل داشتن لایه‌های متعدد و معماری پیچیده که شامل دروازه‌های حافظه برای یادگیری الگوهای بلندمدت است، به‌عنوان یک مدل "یادگیری عمیق" طبقه‌بندی می‌شود. این مدل برای رفع مشکلات موجود در شبکه‌های عصبی بازگشتی با اضافه نمودن تعاملات بیشتر در هر سلول ارائه شده‌است. روش LSTM به دلیل به‌خاطر سپردن اطلاعات برای طولانی مدت روشی خاص می‌باشد که از چهار لایه تعاملی ضروری که روش‌های ارتباطی مختلفی با یکدیگر دارند، تشکیل شده‌است. همچنین این روش از یک بلوک حافظه تشکیل شده‌است. که این بلوک‌ها سلول گفته می‌شود. به‌طوری که اطلاعات در یک سلول ذخیره شده و توسط دروازه‌های کنترل به سلول‌های دیگر منتقل می‌شود. با کمک این دروازه‌ها تجزیه و تحلیل اطلاعات به‌سادگی و با دقت انجام می‌شود. واحدهای LSTM یا بلوک‌ها بخشی از شبکه عصبی تکرار شونده هستند. شبکه عصبی تکرار شونده برای استفاده از فرایندهای حافظه مصنوعی که به این الگوریتم‌های هوش مصنوعی برای شبیه‌سازی اندیشه انسان کمک می‌کنند می‌باشد. بنابراین مزیت استفاده از LSTM این است که قادر است فرایندهای یادگیری متوالی نظارت‌شده را اجرا کند [15].

۲-۵- شبکه‌های عصبی مصنوعی از نوع پرسپترون چندلایه (MLP)

شبکه‌های عصبی مصنوعی از نوع پرسپترون چندلایه (Multi-Layer Perceptron - MLP) یکی از پرکاربردترین معماری‌های یادگیری ماشین نظارت‌شده هستند که در مسائل طبقه‌بندی و رگرسیون، از جمله مدل‌سازی هیدرولوژیکی استفاده می‌شوند. MLP به دلیل توانایی در مدل‌سازی روابط غیرخطی و پیچیده بین متغیرهای ورودی و خروجی، به‌ویژه در داده‌های چندبعدی، ابزار قدرتمندی در تحلیل داده‌های اقلیمی و هیدرولوژیکی محسوب می‌شود. MLP می‌تواند داده‌های چندمتغیره (مانند متغیرهای اقلیمی) را به‌خوبی مدیریت کند و با تنظیم مناسب، MLP می‌تواند الگوهای عمومی را از داده‌ها استخراج کند. مطالعات نشان داده‌اند که MLP در مقایسه با روش‌های آماری سنتی، مانند رگرسیون خطی، دقت بالاتری در پیش‌بینی‌های اقلیمی ارائه می‌دهد.

در فرآیند عملکرد مدل MLP ابتدا داده‌های ورودی (مانند دما، بارش، رطوبت، سرعت باد) نرمال‌سازی می‌شوند تا مقیاس‌های متفاوت تأثیر منفی بر یادگیری نگذارند. تعداد لایه‌های مخفی، تعداد نورون‌ها در هر لایه و تابع فعال‌سازی انتخاب می‌شوند. داده‌های آموزشی برای تنظیم وزن‌ها و بایاس‌ها از طریق پس‌انتشار خطا استفاده می‌شوند. در گام آخر، مدل آموزش‌دیده برای پیش‌بینی داده‌های جدید استفاده شده و با معیارهایی عملکرد ارزیابی می‌شود [16].

۲-۶- طراحی سناریوهای ورودی مدل‌های یادگیری ماشین

در این پژوهش، به‌منظور بررسی تأثیر ترکیب‌های مختلف متغیرهای اقلیمی بر دقت پیش‌بینی بارش ماه آینده، ۱۴ سناریوی ورودی با استفاده از متغیرهای بارش یک ماه قبل (P-1)، بارش دو ماه قبل (P-2)، دمای یک ماه قبل (T-1)، رطوبت نسبی یک ماه قبل (RH-1) و سرعت باد یک ماه قبل (U-1) طراحی و مورد ارزیابی قرار گرفتند. این متغیرها به دلیل ارتباط مستقیم و غیرمستقیم با الگوهای بارش انتخاب شدند. انتخاب این متغیرها با توجه به مشخصات آماری آن‌ها (جدول ۱) و تنوع بالای اقلیمی در منطقه مورد مطالعه، انجام شده است.

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

انتخاب ترکیب‌های مختلف ورودی امکان تحلیل نقش هر متغیر و تعاملات آن‌ها در بهبود عملکرد مدل‌های یادگیری ماشین، از جمله MLP، SVM، LSTM و ET را فراهم می‌کند. جدول ۲ سناریوهای مختلف ورودی به مدل‌ها را در مطالعه حاضر نشان می‌دهد و به‌عنوان مبنای تحلیل‌های بعدی برای ارزیابی عملکرد مدل‌ها مورد استفاده قرار گرفته است.

جدول ۲: سناریوهای مختلف برای ترکیب‌های ورودی در مطالعه حاضر

Scenario Number	Inputs combination
1	P-1
2	P-1, P-2
3	P-1, T-1
4	P-1, RH-1
5	P-1, U-1
6	P-1, T-1, RH-1
7	P-1, T-1, U-1
8	P-1, T-1, RH-1, U-1
9	P-1, P-2, T-1
10	P-1, P-2, RH-1
11	P-1, P-2, U-1
12	P-1, P-2, T-1, RH-1
13	P-1, P-2, T-1, U-1
14	P-1, P-2, T-1, RH-1, U-1

۷-۲- معیارهای ارزیابی عملکرد مدل‌های یادگیری ماشین

برای ارزیابی عملکرد مدل‌های یادگیری ماشین مورد استفاده در این پژوهش، از سه معیار آماری پرکاربرد شامل ضریب همبستگی پیرسون (r)، جذر میانگین مربعات خطاها (Root Mean Square Error - RMSE)، و ضریب نش-ساتکلیف (Nash-) (Sutcliffe Efficiency - NSE) استفاده شد. این معیارها به دلیل توانایی در سنجش دقت پیش‌بینی‌ها و تطابق بین مقادیر مشاهداتی و شبیه‌سازی‌شده، به‌طور گسترده در مطالعات هیدرولوژیکی و اقلیمی مورد استفاده قرار می‌گیرند. این معیارها در روابط ۱ تا ۳ ارائه شده‌اند.

$$r = \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \quad (1)$$

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

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{(O_i - \bar{O}_i)^2} \quad (3)$$

در این روابط، پارامتر $O(i)$ مقادیر مشاهداتی (واقعی)، $P(i)$ مقادیر شبیه‌سازی شده، \bar{O} میانگین مقادیر مشاهداتی، \bar{P} میانگین مقادیر شبیه‌سازی شده و N تعداد مشاهدات است.

مقدار r بین -۱ و +۱ قرار می‌گیرد، که مقدار +۱ نشان‌دهنده همبستگی خطی کامل و مثبت مقادیر پیش‌بینی‌شده با مشاهدات و عملکرد قوی مدل، مقدار -۱ همبستگی خطی کامل و منفی و $r = 0$ نشان‌دهنده عدم وجود همبستگی بین مقادیر

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

پیش‌بینی شده با مشاهدات است. RMSE مقیاسی از خطای پیش‌بینی است و مقادیر کمتر آن نشان‌دهنده دقت بالاتر مدل است. این معیار به واحد داده‌های مشاهداتی وابسته است و امکان مقایسه مستقیم بین مدل‌ها را در یک مجموعه داده فراهم می‌کند. NSE مقیاسی از کارایی مدل نسبت به یک مدل پایه (میانگین مشاهدات) است. مقادیر NSE نزدیک به ۱ نشان‌دهنده عملکرد عالی مدل و مقادیر منفی نشان‌دهنده عملکرد ضعیف‌تر از میانگین مشاهدات است.

۳- نتایج و بحث

۳-۱- ارزیابی عملکرد مدل‌های یادگیری ماشین در پیش‌بینی بارش ماهانه

برای ارزیابی عملکرد مدل‌های یادگیری ماشین، شامل حافظه طولانی کوتاه‌مدت (Long Short Term Memory - LSTM)، درختان اضافی (Extra Trees - ET)، ماشین بردار پشتیبان (Support Vector Machine - SVM)، و پرسپترون چندلایه (Multi-Layer Perceptron - MLP)، در پیش‌بینی بارش‌های ماهانه در دشت مغان، از داده‌های بلندمدت هواشناسی استفاده شد که شامل ۳۶۰ رکورد ماهانه بود. این داده‌ها به دو بخش تقسیم گردید: ۲۵۲ رکورد اول برای آموزش مدل‌ها و ۱۰۸ رکورد آخر برای تست. ارزیابی عملکرد بر اساس معیارهای استاندارد شامل معیارهای آماری ضریب نش-ساتکلیف (NSE)، جذر میانگین مربعات خطاها (RMSE) و ضریب همبستگی پیرسون (r) انجام شد. در جدول ۳ نتایج ارزیابی عملکرد مدل‌ها در مراحل آموزش و صحت‌سنجی، با در نظر گرفتن سناریوهای مختلف ورودی معرفی شده (در جدول ۲) ارائه شده است. این نتایج بهترین ساختار ورودی برای هر مدل را مشخص کرده و امکان مقایسه عملکرد مدل‌ها را فراهم می‌کند.

نتایج جدول ۳ نشان‌دهنده عملکرد قوی مدل‌های LSTM، ET و SVM در مقایسه با MLP است. در بخش آموزش، مدل LSTM با $RMSE = 8.23$ mm، $NSE = 0.87$ و $r = 0.91$ بهترین عملکرد را نشان داد، که حاکی از تطابق بالای پیش‌بینی‌ها با داده‌های مشاهداتی است. این مدل به‌ویژه در ثبت روابط غیرخطی داده‌های بارش موفق عمل کرده است. مدل ET با $RMSE = 8.56$ mm، $NSE = 0.81$ و $r = 0.87$ در رتبه بعدی قرار گرفت. مدل SVM با $RMSE = 9.06$ mm، $NSE = 0.72$ و $r = 0.84$ عملکرد متوسطی داشت، که ممکن است ناشی از حساسیت این مدل به نویز داده‌ها باشد. در نهایت، مدل MLP با $RMSE = 10.10$ mm، $NSE = 0.63$ و $r = 0.81$ ضعیف‌ترین عملکرد را در بخش آموزش نشان داد.

در بخش صحت‌سنجی، که برای ارزیابی تعمیم‌پذیری مدل‌ها استفاده می‌شود، مدل LSTM با $NSE = 0.84$ ، $RMSE = 8.41$ mm و $r = 0.89$ بهترین عملکرد را در میان مدل‌ها ارائه می‌دهد. این امر می‌تواند به توانایی مدل LSTM در به‌خاطر سپردن اطلاعات برای طولانی مدت و مقاومت نسبی آن در برابر بیش‌برازش مرتبط باشد. مدل‌های ET و SVM نیز با مقادیر NSE به ترتیب ۰/۷۹ و ۰/۷۰ و r به ترتیب ۰/۸۶ و ۰/۸۳ در مرحله صحت‌سنجی، عملکرد قابل‌قبولی دارند، که نشان‌دهنده مقاومت این مدل مبتنی بر درخت (ET)، و مدل‌سازی روابط غیرخطی با استفاده از کرنل‌های مناسب (مانند RBF) و بهینه‌سازی حاشیه حداکثری در مدل SVM و توانایی آنها در برابر بیش‌برازش و مدیریت داده‌های اقلیمی نویری است [14]. در مقابل، مدل MLP با $NSE = 0.62$ ، $RMSE = 10.23$ mm و $r = 0.81$ عملکرد نسبتاً ضعیف‌تری نسبت به سایر مدل‌ها دارد، که ممکن است به حساسیت آن به تنظیم‌های اولیه (مانند تعداد لایه‌ها و نوروها) یا نیاز به داده‌های آموزشی بیشتر مربوط باشد [22].

نکته قابل‌توجه در جدول ۳، درمورد تأثیر سناریوهای ورودی است، جایی‌که سناریو ۲ (بارش یک ماه قبل و بارش دو ماه قبل) به‌عنوان بهترین ساختار ورودی برای همه مدل‌های هوشمند شامل LSTM، ET، SVM و MLP انتخاب شده است. این امر نشان‌دهنده نقش کلیدی بارش ماه‌های قبل در پیش‌بینی بارش ماه آتی در دشت مغان است. عدم انتخاب سناریوهای دیگر مخصوصاً سناریو ۱۴ (ترکیب کامل همه متغیرهای ورودی) به‌عنوان بهترین سناریو برای هیچ‌کدام از مدل‌ها، ممکن است به

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

افزایش پیچیدگی مدل و خطر بیش‌برازش یا حضور نویز در متغیرهایی مثل سرعت باد (U) نسبت داده شود، به‌ویژه با توجه به انحراف معیار پایین این متغیر (۰/۸۶ متر بر ثانیه) که در جدول ۱ ارائه شد.

تمامی مدل‌ها در مرحله آموزش عملکرد بهتری (NSE و r بالاتر، RMSE پایین‌تر) نسبت به مرحله صحت‌سنجی نشان می‌دهند، که امری متداول در مدل‌های یادگیری ماشین است و به تفاوت بین داده‌های آموزشی و آزمایشی مربوط می‌شود. با این حال، کاهش اندک (یا عدم تغییر) در معیارهای عملکرد (مانند NSE و r) در مرحله صحت‌سنجی برای مدل MLP (از ۰/۶۳ به ۰/۶۲ برای NSE و مقدار یکسان ۰/۸۱ برای r) نشان‌دهنده تعمیم‌پذیری بالای این مدل است. در مقابل، کاهش بیشتر نسبت به سایر مدل‌ها در مقدار r برای مدل LSTM (از ۰/۹۱ به ۰/۸۹) ممکن است به حساسیت بیشتر این مدل به تغییرات در داده‌های آزمایشی اشاره داشته باشد.

نتایج جدول ۳ نشان‌دهنده برتری نسبی مدل LSTM با ورودی بارش یک ماه قبل و بارش دو ماه قبل در پیش‌بینی بارش یک ماه آینده در دشت مغان است، که می‌تواند به سیاست‌گذاران و مدیران منابع آب در این منطقه کشاورزی‌محور کمک کند تا استراتژی‌های مؤثری برای مدیریت برنامه‌های آبیاری و مدیریت منابع آب در این منطقه توسعه دهند.

جدول ۳: نتایج ارزیابی مدل‌های هوشمند در مراحل آموزش و صحت‌سنجی

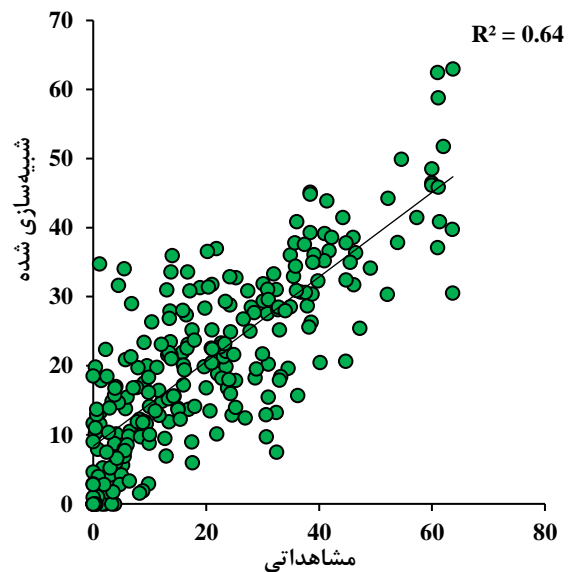
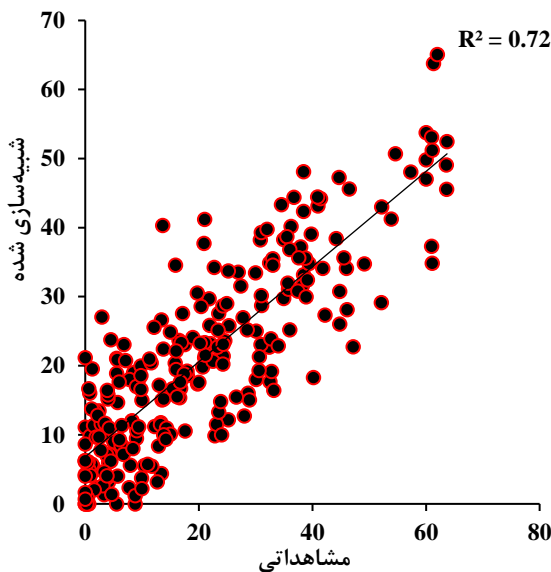
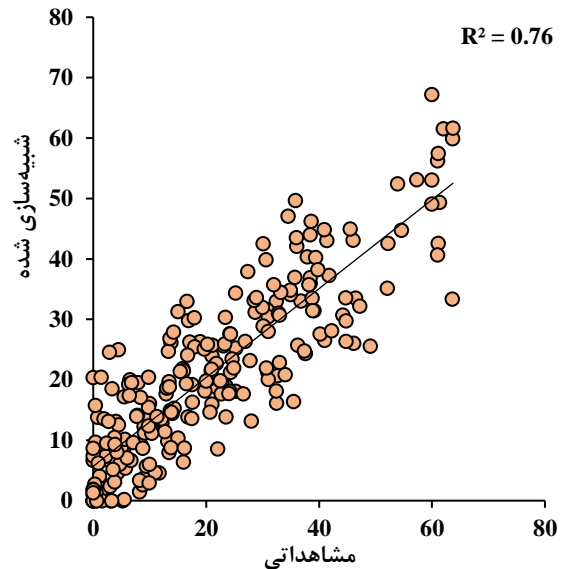
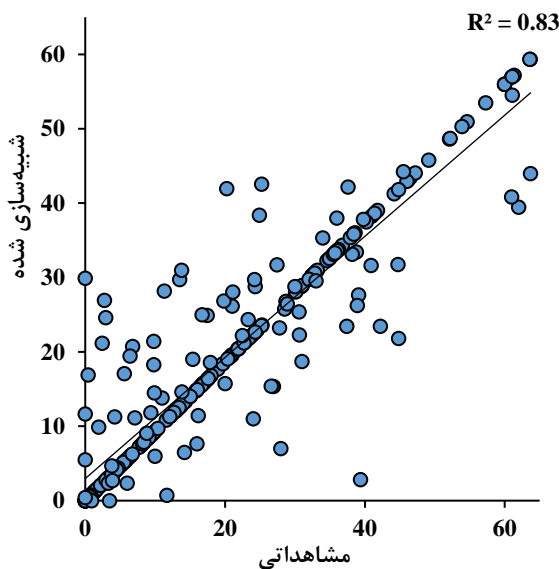
Model	Best scenario	Stage	NSE	RMSE (mm)	r
LSTM	2	Train	0.87	8.23	0.91
		Test	0.84	8.41	0.89
ET	2	Train	0.81	8.56	0.87
		Test	0.79	8.79	0.86
SVM	2	Train	0.72	9.06	0.84
		Test	0.70	9.48	0.83
MLP	2	Train	0.63	10.16	0.81
		Test	0.62	10.23	0.81

علاوه بر ارزیابی عملکرد مدل‌های هوشمند در پیش‌بینی بارش ماهانه، نمودارهای پراکنش مقادیر مشاهداتی در مقابل پیش‌بینی‌شده برای هر مدل به‌صورت جداگانه در بخش‌های آموزش و آزمون نیز ترسیم شد تا رفتار مدل‌ها به‌صورت بصری تحلیل شود. شکل ۴ نمودار پراکنش متقابل مقادیر بارش مشاهداتی و پیش‌بینی‌شده توسط مدل‌های یادگیری ماشین را در مرحله آموزش نشان می‌دهد. این نمودارها امکان ارزیابی بصری و کمی تطابق بین مقادیر مشاهداتی و پیش‌بینی‌شده بارش را فراهم می‌کنند و مبنایی برای سنجش دقت مدل‌ها در پیش‌بینی بارش ماهانه در دشت مغان ارائه می‌دهند.

در بخش آموزش، نمودار پراکنش مدل LSTM نقاطی بسیار نزدیک به خط ایده‌آل ($y = x$) نشان می‌دهد. پراکنش کم در مقادیر متوسط بارش (۱۰ تا ۴۰ میلی‌متر) و انحراف اندک در مقادیر بالا (بیش از ۵۰ میلی‌متر) مشاهده می‌شود، که نشان‌دهنده توانایی بالای مدل در یادگیری الگوهای داده‌های آموزشی است. نمودار پراکنش مدل ET پراکنش بیشتری نسبت به LSTM دارد، به‌ویژه در مقادیر نزدیک به صفر (مثلاً در برخی رویدادهای با بارش صفر میلی‌متر، پیش‌بینی بارش ۲۰/۳۷ یا ۹/۳۷ میلی‌متر مشاهده می‌شود)، که نشان‌دهنده خطای قابل توجه در شناسایی دوره‌های خشک است. نمودار پراکنش مدل SVM پراکنش متوسطی را نشان می‌دهد، درحالی‌که در مقادیر بالا (مثلاً پیش‌بینی بارش ۴۵/۵۵ میلی‌متر برای رویدادهای با بارش ۶۳/۶ میلی‌متر) و مقادیر صفر (مثلاً در برخی رویدادهای با بارش صفر میلی‌متر، پیش‌بینی بارش ۸/۶۵ میلی‌متر) انحرافات

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

بیشتری نشان می‌دهد. نمودار پراکنش مدل MLP پراکنش بیشتری نسبت به LSTM و SVM دارد، به‌ویژه در مقادیر میانی و بالا (مثلاً پیش‌بینی بارش ۳۰/۵۵ میلی‌متر برای رویدادهای با بارش ۶۳/۶ میلی‌متر). این رفتار با مقادیر معیارهای آماری به‌دست آمده در بخش قبلی همخوانی دارد و نشان‌دهنده محدودیت‌های این مدل در ثبت الگوهای پیچیده است.



شکل ۴: نمودار پراکنش مقادیر بارش مشاهداتی و مدل‌سازی شده با استفاده از مدل‌های هوشمند در مرحله آموزش

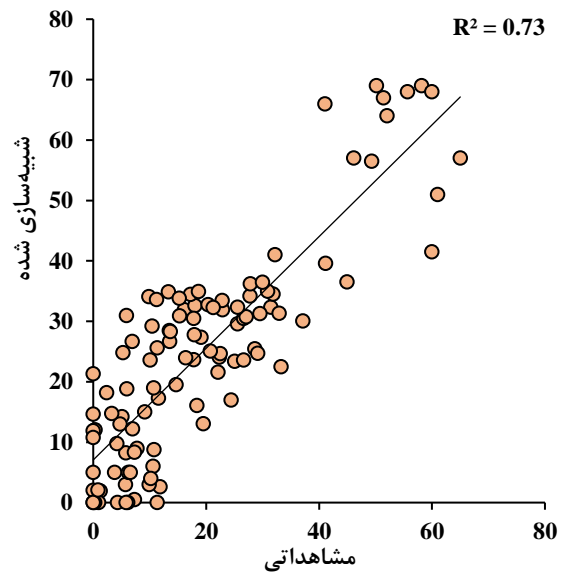
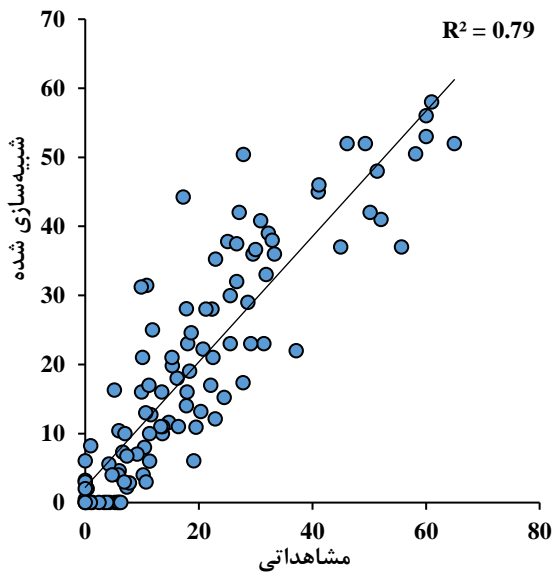
شکل ۵ نمودارهای پراکنش مقادیر بارش مشاهداتی و پیش‌بینی‌شده توسط مدل‌های یادگیری ماشین را در مرحله صحت‌سنجی برای دوره زمانی ژانویه ۲۰۱۴ تا دسامبر ۲۰۲۲ (۱۰۸ ماه) در دشت مغان نشان می‌دهد که نقاط نزدیک به خط $(y = x)$ بیانگر پیش‌بینی‌های دقیق هستند.

نمودار پراکنش مدل LSTM در بخش صحت‌سنجی پراکنش بیشتری نسبت به بخش آموزش نشان می‌دهد، به‌ویژه در مقادیر بالا (مثلاً رویدادهای با بارش ۶۵ میلی‌متر و پیش‌بینی بارش ۵۲ میلی‌متر) و مقادیر صفر (ماه‌های بدون بارش و پیش‌بینی

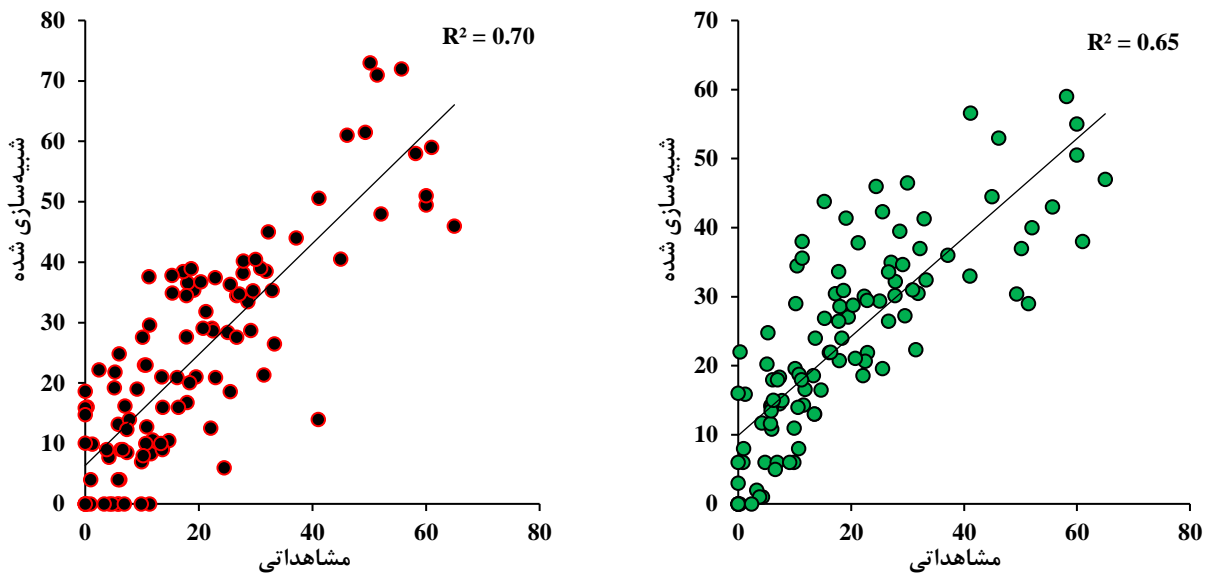
The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

بارش ۲۹/۸۸ میلی متری در این ماهها). این پراکنش نشان دهنده کاهش دقت در تعمیم‌دهی این مدل از بخش آموزش به بخش صحت‌سنجی است. نمودار پراکنش مدل ET در بخش صحت‌سنجی، پراکنش متوسطی دارد، با خطاهای قابل توجه در مقادیر صفر (مثلاً ماه‌های بدون بارش و پیش‌بینی بارش ۱۰/۷۷ یا ۱۴/۶۵ میلی متری در این ماهها) و مقادیر بالا (مثلاً رویدادهای با بارش ۶۵ میلی متر و پیش‌بینی بارش ۵۷ میلی متر در این ماه). نمودار پراکنش مدل SVM تقریباً مشابه مدل ET است، اما خطاها در مقادیر صفر (مثلاً ماه‌های بدون بارش و پیش‌بینی بارش ۱۴/۷۷ یا ۱۸/۶۵ میلی متری در این ماهها) برجسته‌تر هستند. در مقادیر بالا، مدل SVM عملکرد بهتری دارد (مثلاً رویدادهای با بارش ۶۵ میلی متر و پیش‌بینی بارش ۴۶ میلی متر). نمودار پراکنش مدل MLP پراکنش کمتری در مقادیر میانی نشان می‌دهد، اما در مقادیر صفر (مثلاً ماه‌های بدون بارش و پیش‌بینی بارش ۱۳/۴۶ میلی متر) و مقادیر بالا (مثلاً رویدادهای با بارش ۶۵ میلی متر و پیش‌بینی بارش ۴۷ میلی متر) خطاهای این مدل قابل توجه هستند.

این نتایج توانایی مدل‌های یادگیری ماشین، به‌ویژه LSTM، را در ثبت الگوهای بارشی در دشت مغان تأیید می‌کنند و می‌توانند به بهبود پیش‌بینی‌های هیدرولوژیکی، مدیریت پایدار منابع آب و برنامه‌ریزی کاشت، آبیاری و برداشت محصولات کشاورزی کمک کنند.



The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering



شکل ۵: نمودار پراکنش مقادیر بارش مشاهداتی و مدل‌سازی شده با استفاده از مدل‌های هوشمند در مرحله آزمون

به‌طور کلی، تحلیل معیارهای ارزیابی نشان داد که مدل LSTM در بخش آموزش بهترین عملکرد را دارد، اما در بخش تست، مدل MLP تعمیم‌پذیری بهتری ارائه می‌دهد. نمودارهای پراکنش تأیید می‌کنند که همه مدل‌ها در پیش‌بینی مقادیر صفر یا نزدیک به صفر (دوره‌های خشک) با چالش مواجه هستند. مدل LSTM به دلیل ساختار مناسب برای داده‌های زمانی، ابزار مناسبی برای مدل‌سازی بارش در دشت مغان است، اما تنظیم‌های اضافی برای کاهش بیش‌برازش می‌تواند عملکرد آن را در تست بهبود بخشد. این نتایج بر ضرورت استفاده از تکنیک‌های منظم‌سازی و اعتبارسنجی متقابل برای بهبود پیش‌بینی‌های آینده تأکید دارند.

۴- نتیجه‌گیری

تحلیل عملکرد مدل‌های هوشمند یادگیری ماشین شامل شبکه عصبی طولانی کوتاه‌مدت (LSTM)، درختان اضافی (ET)، ماشین بردار پشتیبان (SVM) و شبکه عصبی پرسپترون چندلایه (MLP) در پیش‌بینی بارش‌های ماهانه دشت مغان نشان داد که هر مدل نقاط قوت و ضعف خاص خود را دارد. در بخش آموزش، مدل LSTM به دلیل توانایی بالای خود در ثبت روابط غیرخطی و الگوهای زمانی داده‌های بارش، بهترین تطابق را با مقادیر مشاهداتی ارائه داد. با این حال، در بخش تست، مدل MLP به دلیل تعمیم‌پذیری بهتر، عملکرد مناسب‌تری از خود نشان داد، در حالی که LSTM نشانه‌هایی از بیش‌برازش را بروز داد. مدل‌های ET و SVM نیز عملکرد قابل‌قبولی داشتند، اما در مقایسه با MLP در بخش تست کمی ضعیف‌تر عمل کردند. تحلیل نمودارهای پراکنش نشان داد که همه مدل‌ها در پیش‌بینی دوره‌های خشک (بارش نزدیک به صفر) با چالش‌هایی مواجه هستند و اغلب مقادیر غیرصفر پیش‌بینی می‌کنند، که می‌تواند به پیچیدگی‌های ذاتی داده‌های هواشناسی یا تنظیم ناکافی مدل‌ها مرتبط باشد. این نتایج بر اهمیت استفاده از مدل‌های یادگیری ماشین با ساختار مناسب برای داده‌های زمانی، مانند LSTM، و ضرورت به‌کارگیری تکنیک‌های منظم‌سازی و اعتبارسنجی متقابل برای بهبود دقت پیش‌بینی‌ها تأکید دارد. در نهایت، این مطالعه نشان‌دهنده پتانسیل بالای الگوریتم‌های یادگیری ماشین در مدل‌سازی بارش و حمایت از مدیریت پایدار

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering



منابع آب در مناطق کشاورزی محور مانند دشت مغان است. پیشنهاد می‌شود در تحقیقات آینده، بهینه‌سازی هایپر پارامترها و ترکیب مدل‌ها برای افزایش دقت بررسی شود.

مراجع

1. Allahverdipour, P., & Sattari, M.T. 2023. Comparing the performance of the multiple linear regression classic method and modern data mining methods in annual rainfall modeling (Case study: Ahvaz city). *Water and Soil Management and Modelling* 3(2):125-142. <https://doi.org/10.22098/mmws.2022.11337.1120>
2. Pirone, D., Cimorelli, L., Del Giudice, G., & Pianese, D. 2023. Short-term rainfall forecasting using cumulative precipitation fields from station data: a probabilistic machine learning approach. *Journal of Hydrology*, 617, 128949. <https://doi.org/10.1016/j.jhydrol.2022.128949>
3. Baig, F., Ali, L., Faiz, M.A., Chen, H., & Sherif, M. 2024. How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment?. *Journal of Hydrology*, 633, 131040. <https://doi.org/10.1016/j.jhydrol.2024.131040>
4. Weesakul, U., Kaewprapha, P., Boonyuen, K., & Mark, O. (2018). Deep learning neural network: A machine learning approach for monthly rainfall forecast, case study in eastern region of Thailand. *Engineering and Applied Science Research*, 45(3), 203-211.
5. Sattari, M.T., Bagheri, R., Shirimi, K., & Allahverdipour, P. 2024. Modeling Daily and Monthly Rainfall in Tabriz using Ensemble Learning Models and Decision Tree Regression. *Climate Change Research* 5(18): 31-48. <https://doi.org/10.30488/ccr.2024.433394.1192>
6. Ghosh, S., Gourisaria, M.K., Sahoo, B. & Das, H. 2023. A pragmatic ensemble learning approach for rainfall prediction. *Discover Internet of Things* 3(1):13.
7. Sharafi, M., Samadian Fard, S., & Hashemi, S. 2021. Monthly rainfall Forecasting using genetic programming and support vector machine. *Journal of Rainwater Catchment Systems* 8(4):63-71. <https://dor.isc.ac/dor/20.1001.1.24235970.1399.8.4.2.3>
8. Nazarieh, F., & Ansari, H. 2015. Assessing Intelligent Models in Forecasting Monthly Rainfall by Means of Teleconnection Patterns (Case Study: Khorasan Razavi Province). *Water and Soil* 29(2):274-283. <https://doi.org/10.22067/jsw.v0i0.25357>
9. Ghasemiyeh, H., Bazrafshan, O., & Bakhshayesh manesh, K. 2017. Artificial Neural Network for Monthly Rainfall Forecasting Using Teleconnection Patterns (Case Study: Central Plateau Basin of Iran). *Journal of the Earth and Space Physics* 43(2):405-418. <https://doi.org/10.22059/jesphys.2017.58913>
10. Yang, J., Xiang, Y., Sun, J., & Xu, X. 2022. Multi-model ensemble prediction of summer precipitation in China based on machine learning algorithms. *Atmosphere*, 13(9), 1424. <https://doi.org/10.3390/atmos13091424>
11. Roushangar, K., & Ghasempour, R. 2020. Monthly precipitation prediction improving using the integrated model based on kernel-wavelet and complementary ensemble empirical mode decomposition. *Amirkabir Journal of Civil Engineering* 52(10):2649-2660. <https://doi.org/10.22060/ceej.2019.16043.6109>
12. Vapnik V. 2013. *The nature of statistical learning theory*. Springer science & business media.
13. Pai P F, & Hong W C. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrol. Process* 21: 819-827.
14. Geurts P, Ernst D, & Wehenkel L. 2006. Extremely randomized trees. *Machine learning* 63(1):3-42.
15. Yin L. 2023. Forecasting Sector Rotation of A-share Market Using LSTM and Random Forest. *Advances in Economics, Management and Political Sciences*. *Adv. Econ. Manage. Polit. Sci* 49:109-123.
16. Jain S K, Das A, Srivastava D K. 1999. Application of ANN for reservoir inflow prediction and operation. *J. Water Resour. Plann. Manage* 125:263-271.

The First National Conference on Hydro-Informatics and Artificial Intelligence in Water Engineering

Modeling Monthly Precipitation Using Intelligent Machine Learning and Deep Learning Models (Case Study: Moghan Plain)

Pouya Allahverdipour^{1*} , Yagob Dinpashoh² 

¹ PhD Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. a.pouya1996@gmail.com

² Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. dinpashoh@yahoo.com

Abstract

Precipitation, as a critical component of the hydrological cycle, plays a pivotal role in water resource management and agriculture in sensitive regions such as the Moghan Plain. This study aimed to model monthly precipitation in the Moghan Plain using machine learning and deep learning algorithms, including Long Short-Term Memory (LSTM), Extra Trees (ET), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). Long-term meteorological data, including precipitation, average temperature, average humidity, and average wind speed, comprising 360 monthly records from 1993 to 2022, were collected and divided into two subsets: 252 records for training (70%) and 108 records for testing (30%). Model performance was evaluated using correlation coefficient (r), Root Mean Square Error (RMSE), and Nash-Sutcliffe efficiency (NSE). In the training phase, the LSTM model exhibited the best performance with RMSE = 8.23 mm, NSE = 0.87, and $r = 0.91$. In the validation phase, LSTM also outperformed others with RMSE = 8.41 mm, NSE = 0.84, and $r = 0.89$, while ET (RMSE = 8.79 mm, NSE = 0.79, $r = 0.86$), SVM (RMSE = 9.48 mm, NSE = 0.70, $r = 0.83$), and MLP (RMSE = 10.23 mm, NSE = 0.62, $r = 0.81$) showed weaker performance. Scatter plot analysis revealed that all models struggled with predicting near-zero precipitation values. The results highlight LSTM's superior performance in both training and testing phases, underscoring the high potential of deep learning in precipitation forecasting and supporting sustainable water resource management.

Keywords: Support Vector Machine, Extra Trees, Neural Network, Water Resources, Artificial Intelligence, Machine Learning