پیش‌بینی مقادیر حداکثر بارش روزانه با استفاده از سیستم‌های هوشمند و مقایسه آن با مدل درختی M5؛ مطالعه موردی ایستگاه‌های اهر و جلفا

محمدتقی ستاری[۱] و فرناز نهرین[۲]

تاریخ دریافت ۱۳۹۱/۱۱/۱۸
تاریخ پذیرش: ۱۳۹۲/۰۸/۲۸

چکیده

بارش یکی از مهمترین اجزا چرخه آب بوده و در سنجش خصوصیات اقلیمی هر منطقه ای نقش بسیار مهمی ایفا می‌کند. پیش‌بینی مقادیر بارش حداکثر روزانه در ماه برای اهداف مختلفی نظیر برآورد سیلاب، رواناب، رسوب، برنامه‌ریزی آبیاری و مدیریت حوضه‌های آبریز دارای اهمیت زیادی است. پیش‌بینی بارش در هر منطقه‌ای نیازمند وجود داده‌های دقیق اندازه‌گیری شده از قبیل رطوبت، دما، فشار، سرعت باد و غیره می‌باشد. محدودیت‌هایی از قبیل نبود اطلاعات کافی در مورد مقدار بارش در مقیاس‌های زمانی و مکانی و همچنین پیچیدگی روابط بین پارامترهای هواشناسی مرتبط با بارش، محاسبه این پارامتر با استفاده از روش‌های معمول را غیردقیق و غیرقابل اعتماد می‌کند. در این تحقیق پارامترهای هواشناسی ایستگاه‌های اهر و جلفا در استان آذربایجان شرقی، به عنوان ورودی مدل‌های هوشمند شبکه‌های عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 تعریف گردید و برای نتایج بدست آمده از این سه مدل دو آماره R و RMSE محاسبه گردید. در دو ایستگاه اهر و جلفا روش برنامه‌ریزی ژنتیک به ترتیب با (R=0.88) و (RMSE=3.32) و (R=0.87) و (RMSE=3.79) بهترین نتیجه را نشان دادند. در حالت کلی می‌توان گفت که هر سه روش مذکور ضمن رقابت با یکدیگر نتایج نسبتا دقیقی را جهت پیش‌بینی حداکثر بارش روزانه در ماه مورد نظر در منطقه ارائه می‌کنند ولی به دلیل ارائه روابط خطی ساده و قابل فهم توسط مدل درختی M5، این روش می‌تواند به عنوان روشی کاربردی و جایگزین برای محاسبه حداکثر بارش روزانه در ماه مورد توجه قرار گیرد.

واژه‌های کلیدی: اهر و جلفا، برنامه‌ریزی ژنتیک، پیش‌بینی بارش حداکثر روزانه، شبکه‌های عصبی مصنوعی، مدل درختی M5.

[۱]- گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
[۲]- دانشجوی کارشناسی ارشد سازه های آبی گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
(*- نویسنده مسئول: EMAIL: mtsattar@yahoo.com)

مقدمه

بارش، یکی از مهمترین داده‌های ورودی به سیستم هیدرولوژیکی است و به عنوان یکی از مهمترین پارامترهای اقلیمی در زمینه مدیریت منابع آب محسوب می‌شود. پیش‌بینی این پدیده برای اهداف مختلفی نظیر برآورد سیلاب، رواناب، رسوب، مدیریت حوضه آبخیز، کشاورزی، برنامه‌ریزی آبیاری و امثال آن دارای اهمیت زیادی است. پیش‌بینی بارش در مدیریت و هشدار سیل نیز نقش مهمی بر عهده دارد. به عنوان مثال در زمینه سیلاب، به منظور کاهش خسارات ناشی از سیل با اطلاع از میزان بارندگی می‌توان امکان وقوع سیل را در منطقه پیش‌بینی و اقدامات لازم را به عمل آورد (فلاحی و همکاران، ۱۳۹۰). همچنین با اطلاع از میزان بارندگی موثر می‌توان نیاز خالص آبیاری را جهت تنظیم برنامه آبیاری محاسبه کرد. با توجه به پیشرفت‌های تکنولوژی اگرچه این پیش‌بینی در دامنه زمان و مکان امری محال نیست، ولی از پیچیدگی زیادی برخوردار است. چرا که این رفتار پیچیده و غیرخطی بوده و خود تابع عوامل بسیاری از جمله فشار، دما، رطوبت و ... می‌باشد. امروزه سیستم‌های هوشمند همچون روش شبکه عصبی مصنوعی و برنامه‌ریزی ژنتیک در مدل‌سازی فرایندهای هیدرولوژیکی و مهندسی آب مورد استفاده قرار گرفته‌اند. در سال‌های اخیر نیز مدل‌های درختی بعنوان یکی از تکنیک‌های داده کاوی در پیش‌بینی پارامترهای هیدرولوژیکی و هواشناسی مورد توجه قرار گرفته‌اند. روش‌های مذکور به عنوان روش‌هایی معتبر و شناخته شده در مدل‌سازی فرآیندهای پیچیده غیرخطی قادرند با استفاده از داده‌های هواشناسی ثبت شده در ایستگاه‌های هواشناسی تخمین دقیقتری از پارامترهای هواشناسی منطقه ارائه دهند. شبکه‌های عصبی مصنوعی جزو روش‌های جعبه سیاه بوده و بدون در نظر گرفتن فرایند فیزیکی حاکم بر سیستم و تنها با کمک داده‌های ورودی و خروجی قادر به ایجاد رابطه بین آن دو می‌باشد. در ایران استفاده از این مدل‌های هوشمند جهت پیش‌بینی بارش با اقبال خوبی مواجه بوده است. سینگ و همکاران (۲۰۱۰) برای تخمین میانگین سیلاب سالانه در حوضه‌های آبریز هند از مدل درختی M5 استفاده نمودند. خلیلی و همکاران (۱۳۸۷) با استفاده از شبکه‌های عصبی پرسپترون چندلایه[3] میزان بارش را با دقت قابل قبولی در ایستگاه سینوپتیک مشهد پیش‌بینی کردند. قلی‌زاده و دارند (۱۳۸۸) میزان بارش ماهانه را در تهران با استفاده از شبکه عصبی مصنوعی تخمین زدند. فاتحی مرج و مهدیان (۱۳۸۸) ارتباط بین پدیده انسو[4] را با بارش در حوضه دریاچه ارومیه بررسی و برای پیش‌بینی بارش پاییزه در دو ایستگاه تبریز و ارومیه از مدل شبکه‌های عصبی مصنوعی استفاده کردند. نتایج آن‌ها نشان داد که مدل غیرخطی با استفاده از شاخص‌های اقلیمی مورد استفاده، بارش پاییزه را با دقت بیشتری پیش‌بینی می کند. فلاح قالهری و همکاران (۱۳۸۸) کارکرد دو مدل سیستم استنباط فازی و شبکه عصبی مصنوعی در پیش‌بینی بارش در منطقه خراسان بزرگ را مقایسه و نشان دادند که شبکه عصبی مصنوعی در پیش‌بینی بارش موفق تر از مدل فازی می‌باشد. فلاح قالهری و خوشحال (۱۳۸۸) میزان بارش را در استان خراسان رضوی با استفاده از شبکه‌های عصبی مصنوعی پیش‌بینی کرده و به این نتیجه رسیدند که شبکه عصبی مصنوعی در تمامی سال‌ها می‌تواند بارش را با دقت قابل قبولی پیش‌بینی کند. حلبیان (۱۳۸۸) مقادیر بارش ماهانه یزد را با استفاده از شبکه عصبی مصنوعی پیش‌بینی کردند. ترافالیس و همکاران (۲۰۰۵) از روش شبکه عصبی مصنوعی و چند روش دیگر برای تخمین بارش استفاده کردند که در نهایت روش رگرسیون بردار پشتیان (SVR) را بعنوان روشی که بهترین نتیجه را می دهد معرفی کردند. هاکورتا (۲۰۰۸) بارش‌های موسمی دراز مدت را برای ۳۶ زیربخش هواشناسی هند با استفاده از مدل شبکه عصبی مصنوعی بررسی کرد. فلاح قالهری و همکاران (۲۰۱۰) بارش بهاره استان خراسان رضوی را با استفاده از سیستم استنتاج فازی- عصبی تطبیقی پیش‌بینی کردند و نشان دادند که سیستم استنتاج فازی- عصبی تطبیقی می تواند بارش را با دقت منطقی پیش بینی کند. سلطانی و همکاران (۱۳۸۹) در خوضه آبریز لیقوان فرایند بارش-رواناب را با برنامه‌ریزی ژنتیک مدل سازی نمودند. ساتیش کومار و

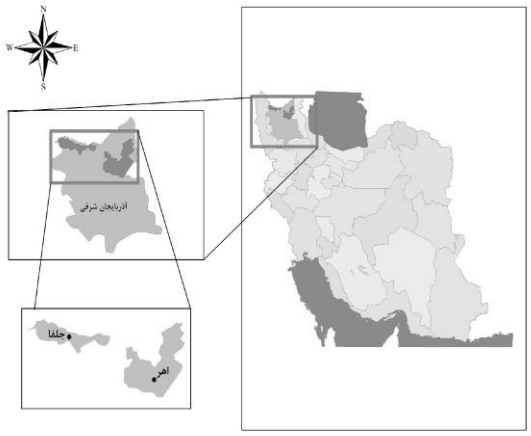
[3]- Multi-Layer Perceptron
[4]- El Nino Southern Oscillation

راجیب (۲۰۱۲) بارش ماهانه در مناطق همگن موسمی هند را با استفاده از برنامه‌ریزی ژنتیک تخمین زدند. فلاحی و همکاران (۱۳۹۰) از مدل رگرسیون درختی به عنوان یک روش کارآمد جهت پیش‌بینی بارندگی در سه ایستگاه هواشناسی بندرانزلی، اراک و قم استفاده کردند که نتایج بیانگر توانایی و دقت بالای این مدل در پیش‌بینی بارش بود. اخیرا مدل‌های درختی در ارتباط با مدلسازی رابطه دبی-اشل (باتا چاریا و سولوماتین ۲۰۰۵)، طبقه‌بندی پوشش زمین (پال، ۲۰۰۶) انتقال رسوب (باتاچاریا و همکاران ۲۰۰۷)، مدل‌سازی تبخیر و تعرق (پال و دئسوال ۲۰۰۹) برآورد باد از امواج (دگا و دئو ۲۰۰۹)، تخمین طیف امواج (ساخار و دئو ۲۰۰۹)، پیش‌بینی ارتفاع موج شاخص (محجوبی و اعتماد شهیدی، ۲۰۰۸، اعتماد شهیدی و محجوبی ۲۰۰۹) و طراحی خاکریزهای سنگی موج شکن (اعتماد شهیدی و بنکدار۲۰۰۹) با موفقیت به کار گرفته شده‌اند. همچنین لاندهه و دیکسیت (۲۰۱۱) مدل درختی **M5** را برای پیش بینی جریان رودخانه در یک روز قبل در دو ایستگاه، رودخانه نارمادا و دیگری در حوضه رودخانه کریشنا در هند به کار بردند. دیتهاکیت و چیناراسری (۲۰۱۲) با استفاده از مدل درختی **M5** اقدام به تخمین ضریب تشتک تبخیر نمودند. هدف از این تحقیق مدل‌سازی حداکثر بارش روزانه شهرستان‌های اهر و جلفا با استفاده از سیستم‌های هوشمندی چون شبکه‌های عصبی مصنوعی و برنامه‌ریزی ژنتیک و مقایسه آن با مدل درختی **M5** می‌باشد. در این مدل‌سازی پارامترهای هواشناسی موثر شامل: دما، باد، رطوبت نسبی، و ... در محاسبه حداکثر بارش روزانه در هر ماه بعنوان ورودی مدل‌ها ارائه شده و پیش‌بینی نسبتا دقیقی از میزان حداکثر بارش روزانه در هر ماه شده است.

مواد و روش‌ها

منطقه مورد مطالعه و داده های مورد استفاده

شهرستان اهر و جلفا از شهرهای مهم استان آذربایجان شرقی می‌باشند. ایستگاه سینوپتیک اهرکه در سال ۱۹۸۵ تاسیس شده است، در طول جغرافیایی ۴۷ درجه و ۴ دقیقه و عرض جغرافیایی ۳۸ درجه و ۲۶ دقیقه و با ارتفاع ۱۳۹۱ متر از سطح دریا قرار گرفته است. ایستگاه سینوپتیک جلفا نیز که در سال ۱۹۸۴ تاسیس شده است، در طول جغرافیایی ۴۵ درجه و ۴۰ دقیقه و عرض جغرافیایی ۳۸ درجه و ۴۵ دقیقه و با ارتفاع ۷۳۶ متر از سطح دریا قرار گرفته است. موقعیت جغرافیایی شهرستان‌های اهر و جلفا در شکل ۱ نشان داده شده است. در این تحقیق از داده‌های حداکثر بارش روزانه ایستگاه هواشناسی اهر و جلفا از ژانویه ۱۹۸۶ لغایت دسامبر ۲۰۰۵ استفاده گردیده که ۸۰ درصد از داده‌ها به عنوان داده‌های آموزشی[5] و ۲۰ درصد داده ها به عنوان داده‌های آزمایشی[6] انتخاب شدند.



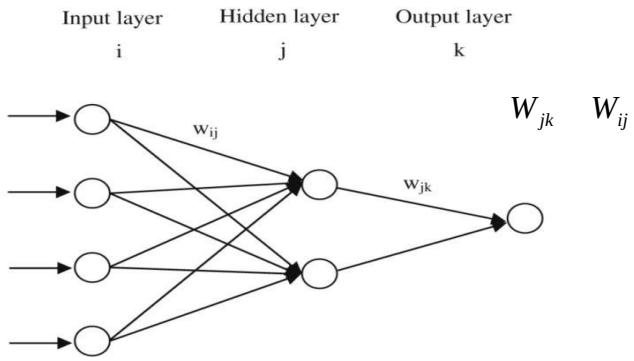
شکل ۱- موقعیت جغرافیایی شهرستان های اهر و جلفا.

[5]- Training Data

[6]- Testing Data

شبکه‌های عصبی مصنوعی پرسپترون‌های چندلایه (MLP)

شبکه‌های عصبی مصنوعی[1] جزو سیستم‌های دینامیکی می‌باشند که با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کنند. به همین سبب به این سیستم‌ها هوشمند گفته می‌شود زیرا که بر اساس محاسبات روی داده‌های مشاهداتی قوانین کلی حاکم بر پدیده‌ها را کشف نموده و بر اساس آنها به پیش‌بینی‌هایی دست می‌زنند. مدل‌های اساسا" فیزیکی به علت سادگی اغلب توسط الگوریتم ANNs قابل مدل‌سازی می‌باشند. مدل‌های ANNs اغلب توسط توپولوژی شبکه، خواص گره‌ها و قوانین آموزش و یادگیری مشخص می‌شوند. در میان الگوهای متفاوت ANNs، شبکه‌های پیش سو[1] نسبت به سایر الگوها از محبوبیت بیشتری برخوردار می‌باشد. ساختار شبکه شامل لایه‌هایی از المان‌های پردازشی موازی بوده که اصطلاحا" به آن نرون گفته می‌شود که در آن هر لایه توسط اتصال‌های میانی وزن‌دار به لایه پردازشی متصل می‌شود. شکل ۲ یک شبکه عصبی سه لایه شامل iو j و k با اتصال دهنده‌های میانی وزنی W_{ij} و W_{jk} بین نرون‌ها را نشان می‌دهد. در هنگام شروع آموزش شبکه، وزن‌ها دارای یک مقدار اولیه بوده که یا توسط مقادیری اتفاقی و یا توسط نتایج آزمایشگاهی پیشین، مقداردهی می‌شوند. اساسا" ایجاد ساختار مناسب شبکه در یک مسئله، توسط سه مرحله تثبیت ساختار، آموزش شبکه و کنترل شبکه انجام می‌گیرد (شریف زاک و بارانی، ۱۳۹۰). در این تحقیق برای مدل‌سازی شبکه عصبی مصنوعی از نرم‌افزار NeuroSolutions 5 استفاده گردیده است.

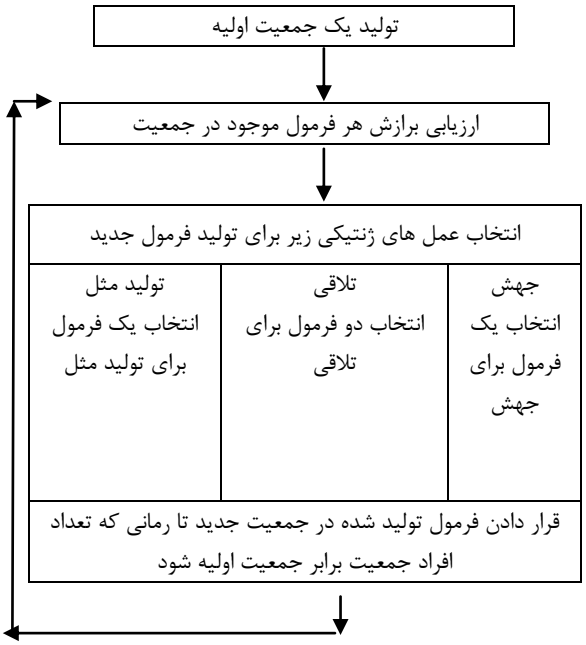


شکل ۲- ساختار یک شبکه عصبی مصنوعی سه لایه.

برنامه ریزی ژنتیک

برنامه ریزی ژنتیک تعمیم یافته روش الگوریتم ژنتیک می‌باشد که برای اولین بار توسط جان کزا (۱۹۹۲) بر اساس تئوری داروین ارائه شد. برنامه ریزی ژنتیک یک تکنیک برنامه‌ریزی خودکار می‌باشد که راه حل مسئله را با استفاده از برنامه‌کامپیوتری ارائه می‌کند. در این روش در ابتدای فرآیند، هیچگونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه سازی ساختار مدل و مؤلفه های آن می باشد. برنامه‌ریزی ژنتیک بر خلاف الگوریتم ژنتیک روی ساختار درختی فرمول ها به جای سلسله ارقامِ دودوئی عمل می‌کند. ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول ها) و ترمینال‌ها (متغیّرهای مسئله و اعداد ثابت) تشکیل می‌شوند. قبل از مراحل اجرایی برنامه ریزی ژنتیک گام های مقدماتی زیر باید توسط کاربر تعیین شوند. ۱. مجموعه ترمینال ها (متغیّر های مسئله، اعداد ثابت تصادفی) ۲. مجموعه عملگرهای ریاضی مورد

استفاده در فرمول ها ۳. انتخاب تابع برازش مناسب (RRSE[۷], MSE[۸], R[۹],...)، برای سنجش برازش فرمول ها ۴. تعیین پارامترهای کنترل کننده اجرای برنامه (اندازه جمعیت، احتمال مربوط به بکار گیری اعمال ژنتیکی و جزئیات دیگر مربوط به اجرای برنامه) ۵. معیار پایان و ارائه نتایج اجرای برنامه (مثل تعداد تولید جمعیت جدید، تعیین یک مقدار مشخص برای برازش فرمول ها که اگر میزان برازش برابر یا بیشتر از آن مقدار شد، اجرا متوقف شود) (سلطانی و همکاران، ۱۳۸۹). فرآیند اجرایی گام به گام برنامه ریزی ژنتیک به صورت مراحل زیر است: ۱. تولید یک جمعیت اولیه از فرمول ها که این فرمول ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول ها) و ترمینال‌ها (متغیّرهای مسئله و اعداد ثابت) ایجاد می‌شوند. ۲. هر یک از افراد جمعیت مذکور با استفاده از توابع برازش مورد ارزیابی قرار می گیرند. ۳. تولید یک جمعیت جدید از فرمول ها، بدین منظور مراحل زیر برای تولید یک جمعیت جدید دنبال می شود: الف) یکی از عمل های ژنتیکی تلاقی[۱۰]، جهش[۱۱] و تولیدمثل[۱۲] انتخاب می شود (این سه عمل ژنتیکی، مهمترین عمل های ژنتیکی مورد استفاده در برنامه ریزی ژنتیک می‌باشند. عمل های دیگری مثل اصلاح ساختار و... نیز با احتمال کمتر مورد استفاده قرار می‌گیرند). ب) تعداد مناسبی از افراد جمعیت حاضر انتخاب می شوند (انتخاب فرد یا افرادی از جمعیت مذکور به صورت احتمالاتی می‌باشد که در این انتخاب احتمالاتی منفردهای با برازش بهتر به منفردهای نامرغوب ترجیح داده می‌شوند و این بدان معنی نیست که حتماً منفردهای نامرغوب حذف می‌شوند). ج) از عمل ژنتیکی انتخاب شده برای تولید فرزند (فرمول جدید) استفاده می شود. د) فرزند (فرمول جدید) تولید شده در یک جمعیت جدید وارد می شود. ه) مدل مورد نظر با استفاده از تابع برازش مورد ارزیابی واقع می شود. ۴. گام سوم تا نیل به حداکثر تعداد تولید، تکرار خواهد شد (سلطانی و همکاران، ۱۳۸۹). شکل کلی گام های اجرایی برنامه ریزی ژنتیک درشکل ۳ نشان داده شده است. در این تحقیق برای اعمال روش برنامه ریزی ژنتیک روی داده ها از نرم افزار GeneXproTools 4.0 استفاده شده است.



[7]- Root relative Square Error
[8]- Mean Square Error
[9]- Correlation oefficient
[10]- Crossover
[11]- Mutation
[12]- Reproduction

مدل درختی M5

مدل درختی M5 که توسط کوئینلن (۱۹۹۲) کشف شده و تئوری او به صورت یک روش توسعه یافت و مدل وی در سال ۱۹۹۷ توسط ونگ و ویتن به نام M5 نامیده شد. این مدل زیر مجموعه‌ای از روش‌های یادگیری ماشینی و داده کاوی است. داده کاوی به فرآیند جست و جو و کشف مدل‌های گوناگون، مختصرسازی‌ها و اخذ مقادیر از مجموعه‌ای از مقادیر معلوم اطلاق می‌گردد (علیخان‌زاده ۱۳۸۵). روش‌های داده کاوی برای مجموعه داده‌های بزرگ با متغیرهای زیاد ساخته شده‌اند، بنابراین متفاوت از روش‌های آماری قدیمی هستند که برای مجموعه داده‌های کوچک با متغیرهای اندک طراحی شده اند. روش‌های بر مبنای درخت[۱۳] یکی از تکنیک‌های داده کاوی است که در این روش‌ها خروجی به صورت یک مدل با سازه درختی با استفاده از داده‌های ورودی و خروجی می‌باشد. الگوریتم M5 رایج‌ترین طبقه‌بندی استفاده شده در خانواده مدل تصمیم‌گیری درختی است. اساسا یک مدل درخت به شکل درخت تصمیم‌گیر با توابع رگرسیون خطی جایگزین مقادیر کلاس ترمینال در برگ‌های درخت ستانده می‌شود. مدل درختی M5 یک الگوریتم پیش‌بینی عددی است و گره‌های درخت با خواص بیشینه خطاهای مورد انتظاری که به عنوان تابعی از انحراف استاندارد پارامترهای خروجی می‌باشد انتخاب می‌شوند. مدل M5 یک مدل درختی برای پیش‌بینی صفات عددی پیوسته است که در آن توابع رگرسیونی خطی در برگ‌های این درخت متظاهر می‌شوند (پال، ۲۰۰۶) که در سال‌های اخیر تحول قابل توجهی در مسایل طبقه‌بندی و پیش‌بینی ایجاد نموده است. مدل‌های درختی چندین مزیت دارند که آن‌ها را به یک روش رگرسیون مناسب برای تجزیه و تحلیل عملکرد تبدیل کرده است. دقت پیش‌بینی مدل درختی با تکنیک‌هایی نظیر شبکه‌های عصبی شناخته شده که می‌تواند بزرگ‌تر از مدل پیش‌بینی رگرسیون خطی نظیر روش کارت باشد لیکن نتایج ساختار درختی و مدل‌های رگرسیون برگ‌ها برای پیشبرد آگاهی از ماهیت و میزان مشکل در عملکرد می‌تواند مورد استفاده قرار گیرد. نتایج مدل درختی برای فهم و شبیه‌سازی آسان هستند و خروجی مدل از دقت بالایی برخوردار است که می‌توان آن را با سایر مدل‌ها مقایسه کرد، مدل درختی می‌تواند در مسایل گوناگون مورد استفاده محققین قرار گیرد (فلاخی و همکاران، ۱۳۹۰). مدل‌های درختی همچنین برای اداره کارآمد مجموعه داده‌های بزرگ با تعداد زیادی از ویژگی‌ها و ابعاد بالا شناخته شده‌اند. یک درخت تصمیم معمولا" از چهار بخش ریشه[۱۴]، شاخه[۱۵]، گره‌ها[۱۶] و برگ‌ها[۱۷] تشکیل شده است که گره‌ها با دایره نشان داده می‌شوند و شاخه‌ها نشان دهنده اتصال بین گره‌ها می‌باشند. درخت تصمیم به منظور سادگی در رسم معمولا" از چپ به راست و یا از بالا به پایین کشیده می‌شود، به طوری که ریشه (گره اول) در بالا قرار می‌گیرد. انتهای یک زنجیره را برگ می‌نامند. هر گره مربوط به یک خصوصیت معین است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند، این بازه‌های مقادیر بخش‌های مختلف مجموعه مقادیر معلوم را برای خصوصیت‌ها به دست دهند. عمل انشعاب توسط یکی از متغیرهای پیش‌بینی کننده انجام می‌پذیرد، بازه‌های انشعاب طوری انتخاب می‌شوند که مجموع مجذور انحراف از میانگین داده‌های هر گره را حداقل کنند (فلاخی و همکاران، ۱۳۹۰). اولین مرحله برای ایجاد یک مدل درختی، استفاده از یک معیار انشعاب است. معیار انشعاب برای الگوریتم M5 بر اساس عملکرد انحراف استاندارد مقادیر هر کلاس و یا طبقه است که در هر گره به دست آمده است. این روش اساس روش‌های طبقه‌بندی است که آنتروپی[۱۸] نامیده می‌شود. آنتروپی می تواند به عنوان معیار میزان آشفتگی و بی‌نظمی یک سیستم تفسیر شود. شکل شماره ۴ نشان می‌دهد که یک ساختار درختی از روش‌های آموزشی مربوط به M5 دامنه پارامتر ورودی ۱ و ۲ می باشد. معیار انشعاب بیانگر میزان خطا در آن گره می‌باشد و مدل حداقل خطای مورد انتظار را به عنوان نتیجه‌ی آزمایش هر صفت در آن گره محاسبه می‌کند. خطای مدل عموما با اندازه‌گیری

[13]- Tree-Based Methods

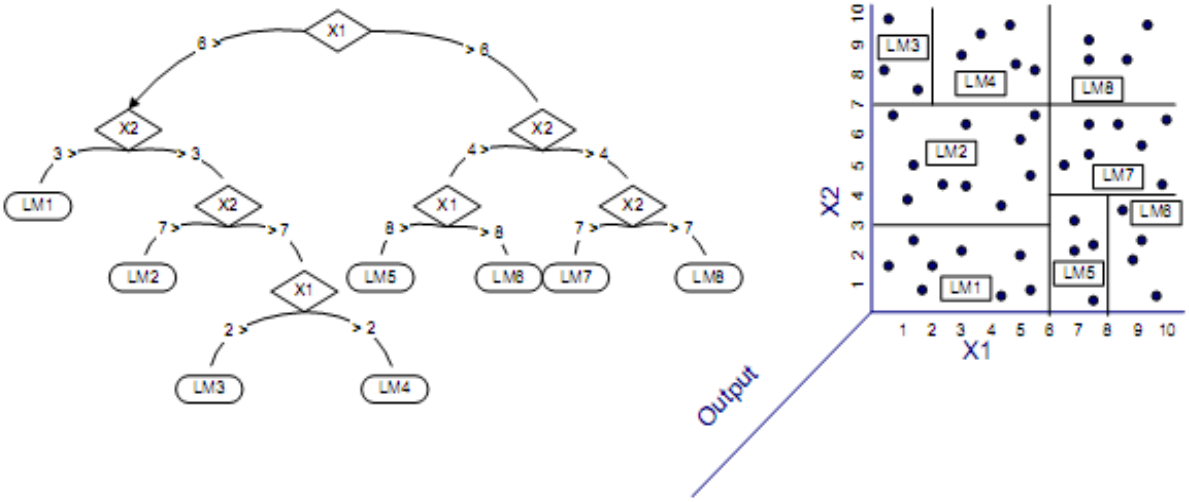
[14]- Root
[15]- Beach
[16]- Nodes
[17]- Leafs
[18]- Entropy

دقت پیش‌بینی مقادیر هدف موارد دیده نشده سنجش می‌شود. فرمول محاسبه کاهش انحراف استاندارد (SDR) به صورت روابط ۱ و ۲ می‌باشد.

$$SDR = Sd(T) - \sum_{i=1}^{N} \frac{|T_i|}{|T|} Sd(T_i) \qquad (۱)$$

$$Sd(T) = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} y_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right)} \qquad (۲)$$

در این رابطه T مجموعه ای از نمونه هاست (موارد) که به هر گره وارد می شود، T_i نشان دهنده ی زیر مجموعه ای از نمونه هاست که I_امین نتیجه تست پتانسیلی را دارند، S_d بیانگر انحراف معیار، y_i مقدار عددی ویژگی هدف نمونه i و N شماره داده ها را نشان می دهد (آلبرگ، ۲۰۱۲).



شکل ۴- نمونه ای از مدل درختی M5.

مدل‌های تصمیم درختی انواع مختلف دارد. هنگامی که خروجی یک درخت، یک مجموعه‌ی گسسته از یک مجموعه مقادیر ممکن است، به آن طبقه‌بندی درختی گفته می‌شود. هنگامی که بتوان خروجی درخت را یک عدد حقیقی در نظر گرفت، آن را رگرسیون درختی می‌نامند. به عبارت دیگر، اگر متغیرهای ورودی به سیستم، عددی[۱۹] باشند، از رگرسیون درختی[۲۰] و اگر مطلق و قیاسی باشند، از طبقه‌بندی درختی[۲۱] استفاده می‌کنیم (فلاخی و همکاران، ۱۳۹۰).

فرآیند انشعاب در هر گره بارها تکرار می‌شود تا به گره پایانی (برگ) برسد که در برگ، مجموع مجذور انحراف از میانگین داده‌ها حدودا" به صفر می‌رسد. با این کار درخت بزرگی توسعه پیدا خواهد کرد. کار با این درخت بزرگ که شاخه‌ها و گره‌های زیادی دارد سخت می‌باشد، بنابراین برای رسیدن به یک درخت بهینه و کارآمد باید شاخه‌های اضافی درخت هرس شود. دو روش برای هرس کردن درخت[۲۲] وجود دارد: ۱) هرس قبل از شکل‌گیری درخت حداکثر[۲۳] ۲) هرس بعد از شکل‌گیری درخت حداکثر[۲۴]. در روش اول

[19]- Numerical
[20]- Regression Tree
[21]- Classification Tree

[22]- Tree Pruning
[23]- Pre-Pruning
[24]- Past-Pruning

فرآیند هرس اجازه نمی‌دهد شاخه‌های اضافی تولید شوند، ولی در روش دوم ابتدا درخت حداکثر تشکیل می‌شود، سپس فرآیند هرس انجام می‌گیرد. درخت بهینه بر اساس حداقل کردن خطای پیش‌بینی انتخاب می‌شود که دو روش برای محاسبه‌ی خطای پیش‌بینی وجود دارد: ۱) آزمون دستگاه مستقل ۲) آزمون صحت‌سنجی. روش اول هنگامی مورد استفاده قرار می‌گیرد که تعداد داده‌های زیادی داشته باشیم، در غیر این صورت از روش دوم استفاده می‌شود (فلاخی و همکاران، ۱۳۹۰) پس از هرس کردن، پروسه‌ی صاف کردن[25] برای جبران گسیختگی‌های تند که به طور اجتناب‌ناپذیر بین مدل‌های خطی هم جوار در برگ‌های درخت هرس شده، مخصوصا" برای مدل‌هایی که از مقادیر کمتری از نمونه‌ها ساخته شده‌اند، انجام می‌شود (باتا چاریا و سولوماتین ۲۰۰۵). در این تحقیق جهت پیش‌بینی حداکثر بارش با استفاده از مدل درختی M5 از نرم افزار WEKA[26] که در دانشگاه Waikato نیوزلند توسعه داده شده، استفاده گردیده است. خصوصیات آماری برای داده‌های مورد استفاده در این تحقیق در جدول ۲ ارائه شده است. در این تحقیق با توجه به ماهیت مسئله ابتدا سناریوهای مختلفی از لایه‌های ورودی، پنهان و خروجی برای شبکه عصبی مصنوعی تعریف گردید و پس از اجرای مدل براساس آماره‌هایی چون RMSE و R، دقت پیش بینی مدل بررسی شد. در این بخش داده ها به دو قسمت آموزش و آزمون تقسیم شدند. جهت مقایسه عملکرد روش شبکه های عصبی با روش برنامه ریزی ژنتیک و مدل درختی M5 اقدام به مدل سازی گردید و نهایتا نتایج مورد مقایسه قرار گرفت. عملکرد مدل های مورد استفاده در این تحقیق بر پایه محاسبه جذر میانگین مربعات خطا (RMSE) و ضریب همبستگی R مورد ارزیابی قرار گرفته است. فرمول‌های محاسبه دو آماره فوق در روابط ۳ و ۴ آمده است.

$$R = \left(\frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2 * \sum_{i=1}^{N}(Y_i - \overline{Y})^2} \right) \qquad (۳)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - X_i)^2}{N}} \qquad (۴)$$

نتایج

برای مقایسه سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 داده‌های مربوط به میانگین دما، فشار سطح ایستگاه، میانگین فشار بخار، میانگین رطوبت نسبی، میانگین سرعت باد، بارش و ساعات آفتابی بعنوان ورودی و حداکثر بارش روزانه بعنوان خروجی همگی در مقیاس زمانی ماهانه وارد سه مدل گردید و مقادیر R و RMSE جهت مقایسه‌ی سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و M5 محاسبه و در جدول ۱ نشان داده شده است.

جدول۱- نتایج حاصل از سه روش متفاوت در پیش بینی حداکثر بارش روزانه در ماه در ایستگاههای اهر و جلفا

روش	R		RMSE	
	اهر	جلفا	اهر	جلفا
ANN	0.86	0.86	12.92	3.78
GP	0.88	0.87	3.32	3.79
M5	0.87	0.86	3.42	3.65

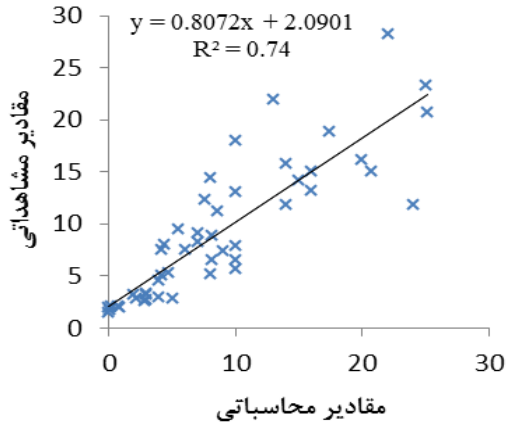
در این قسمت از مدل سازی شبکه های عصبی مصنوعی از تابع تانژانت هیپربولیک بعنوان تابع محرک و قاعده آموزش مومنتوم که باعث همگرا شدن سریع شبکه شده اند، استفاده گردیده است. نتایج مربوط به مقادیر مشاهداتی بارش به دست آمده از مدل شبکه عصبی مصنوعی، برنامه ریزی ژنتیک و مدل درختی M5، در مقابل مقادیر محاسباتی برای پارامتر های ورودی برای ایستگاه اهر به ترتیب در شکل های ۵، ۶ و ۷ و ایستگاه جلفا در شکل‌های

[25] - Smoothing
[26] - Waikato Environment for Knowledge Analysis

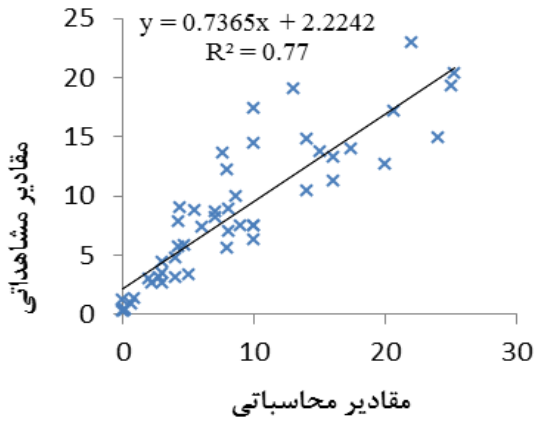
۹، ۱۰ و ۱۱ نشان داده شده است. مقدار بالای R^2 در این شکل‌ها، نشان گر قدرت بالای مدل‌ها در پیش‌بینی بارش در منطقه مورد مطالعه می‌باشد. شکل ۸ و ۱۲ نیز جهت نشان دادن تشابه بالای نتایج حاصل از سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 با نتایج محاسباتی بارش برای داده‌های آزمایش نشان برای شهرستان‌های اهر و جلفا ارائه شده است.

جدول۱- خصوصیت آماری ماهانه داده ها.

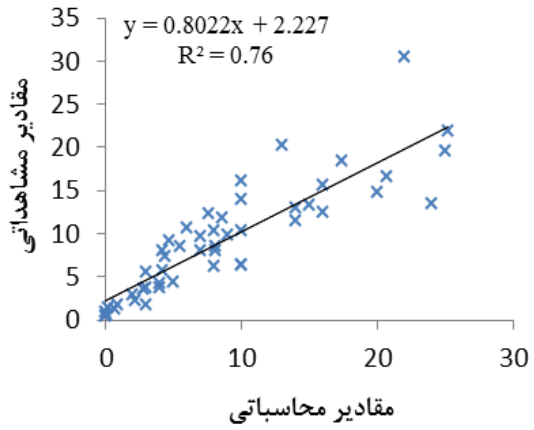
آماره	ایستگاه	میانگین دمای روزانه	فشار سطح ایستگاه	میانگین فشار بخار	میانگین رطوبت نسبی	میانگین سرعت باد	بارش ماهانه	ساعات آفتابی	حداکثر بارش روزانه
میانگین	اهر	10.84	860.84	7.99	60.40	5.64	24.35	214.53	9.23
انحراف استاندارد		8.16	2.18	3.45	8.14	1.62	21.19	67.24	6.79
حداکثر		24.40	867.20	14.57	79.00	12.10	119.6	357.30	34.00
حداقل		-5.70	854.60	3.09	41.00	2.20	0.00	61.60	0.00
میانگین	جلفا	14.88	930.85	9.22	54.66	4.65	16.46	222.20	7.02
انحراف استاندارد		10.34	4.42	3.79	13.39	3.04	17.76	86.05	6.16
حداکثر		31.90	940.70	17.41	86.00	15.20	118.80	370.90	31.10۰
حداقل		-6.70	921.90	2.77	28.00	0.50	0.00	38.20	0.00



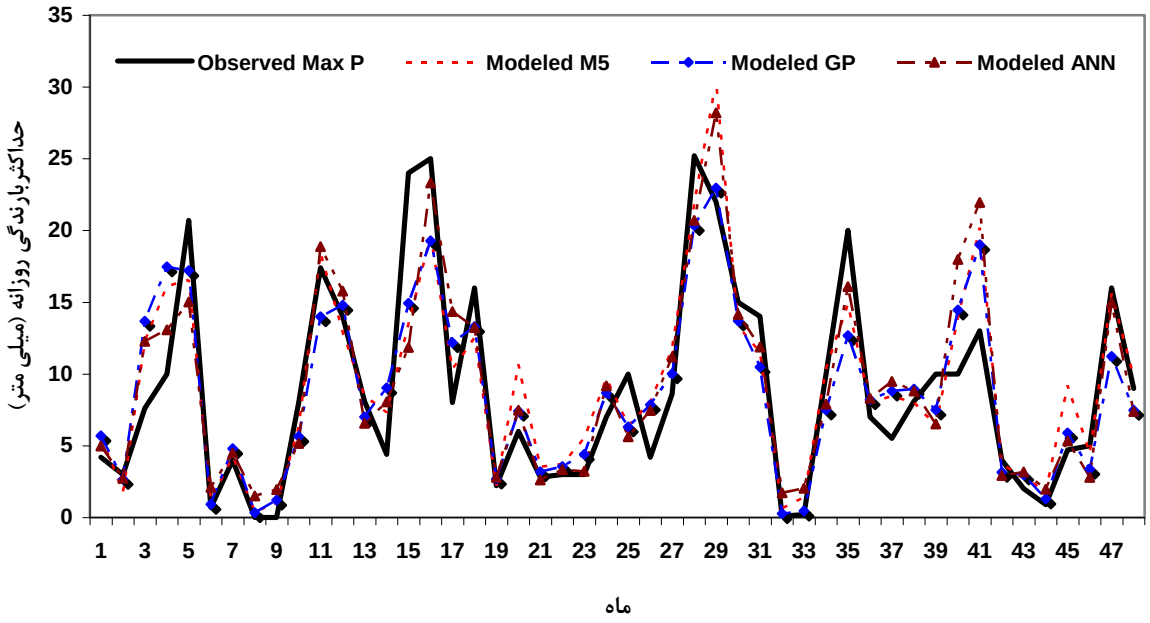
شکل ۵- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل ANN (ایستگاه اهر)



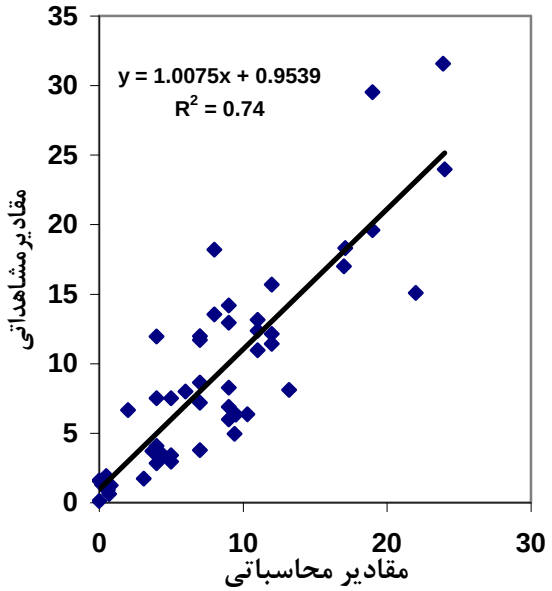
شکل ۶- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل GP (ایستگاه اهر)



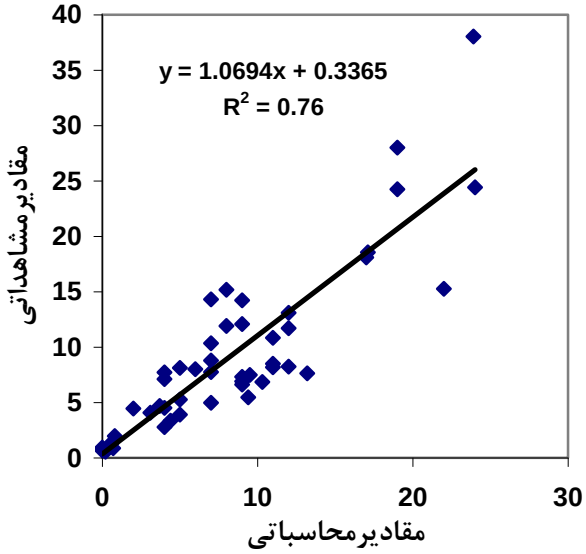
شکل ۷- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل M5 (ایستگاه اهر)



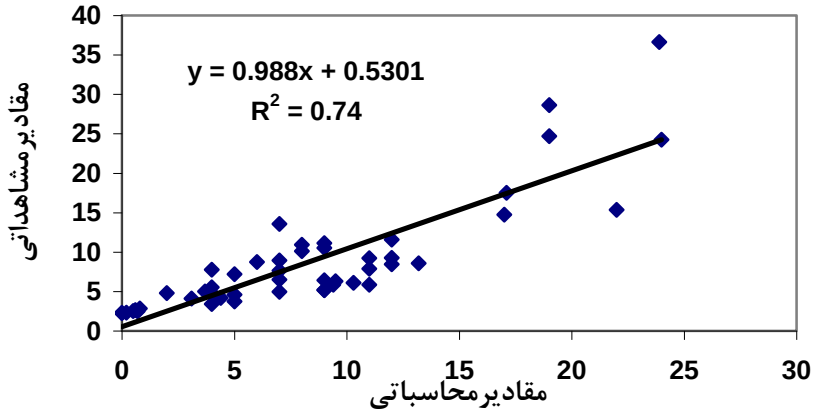
شکل ۸- مقایسه نتایج به دست آمده از روش های مختلف در محاسبه حداکثر بارش روزانه برای ماههای آزمون (ایستگاه اهر)



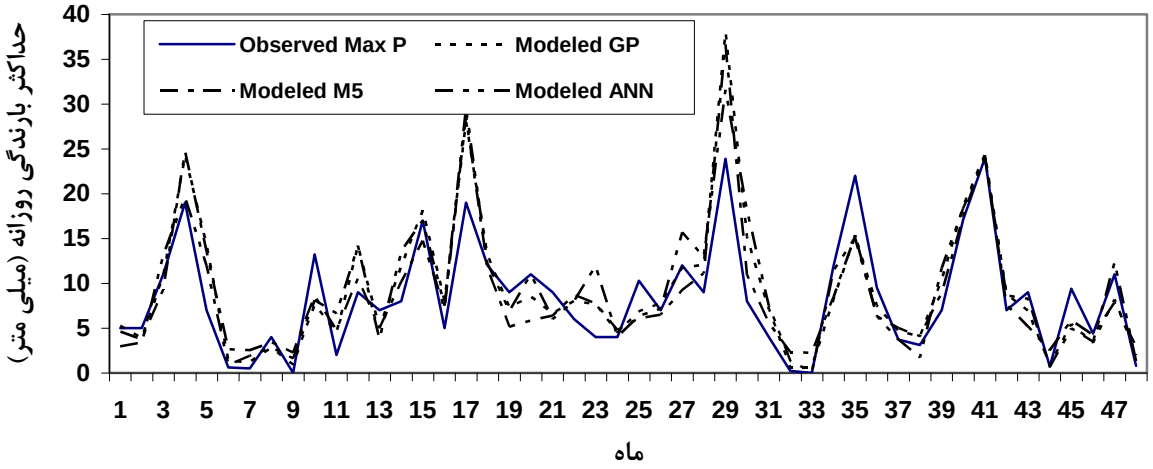
شکل ۹- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل ANN (ایستگاه جلفا)



شکل ۱۰- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل GP (ایستگاه جلفا)

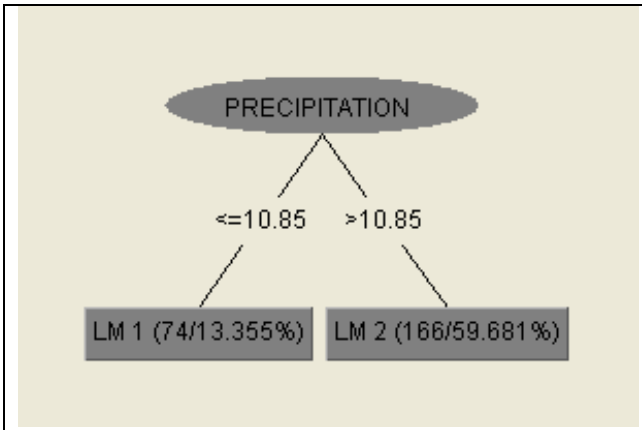


شکل ۱۱- مقادیر محاسباتی و مشاهداتی حداکثر بارش روزانه در مدل M5 (ایستگاه جلفا)



شکل ۱۲- مقایسه نتایج به دست آمده از روش های مختلف در محاسبه حداکثر بارش روزانه برای ماههای آزمون (ایستگاه جلفا).

نتایج حاصل از کاربرد مدل درختی M5 در پیش بینی حداکثر بارش روزانه در شهرستان اهرمنجر به ایجاد ۲ رابطه خطی و در شهرستان جلفا یک رابطه خطی شده است که در شکل های ۱۳ و ۱۴ نشان داده شده است. در شهرستان جلفا بعلت ارائه یک رابطه خطی نمایش درختی مدل ارائه نشده است.



MONTHLY PRECIPITATION <=10.85: LM1
LM num: 1
GDP = 0.0576 * AV TEMP + 0.2662 * QFE PRES - 0.0266 * AV RH + 0.0673 * AV WIND + 0.5926 * PRECIPITATION - 228.5147

MONTHLY PRECIPITATION > 10.85: LM2
LM num: 2
GDP = 0.0044 * AV TEMP + 0.6322 * QFE PRES - 0.1408 * AV RH + 0.0331 * AV WIND + 0.2313 * PRECIPITATION - 531.0787

شکل ۱۳- مدل درختی M5 و روابط خطی جهت پیش بینی حداکثر بارش روزانه (ایستگاه اهر)

روابط خطی مذکور در بازه‌های مشخصی از داده‌های ورودی ارائه شده است. اعداد داخل پرانتز در مدل درختی M5 به ترتیب تعداد مواردی از داده‌ها را که در شرایط هر رابطه خطی صدق می‌کند و درصد انحراف در گره‌های مربوط به هر رابطه خطی ناظر به قانون مذکور را بصورت درصد نشان می‌دهد. بعنوان مثال برای ایستگاه اهر در مدل درختی ارائه شده در شکل ۱۳ چنانچه مقدار بارش در ماه کوچک‌تر یا مساوی ۱۰/۸۵ باشد در آنصورت مقدار حداکثر بارش روزانه در ماه با استفاده از رابطه خطی LM1 و اگر مقدار بارش بزرگتر از ۱۰/۸۵ باشد، حداکثر بارش روزانه از

رابطه خطی LM2 قابل محاسبه خواهد بود. همچنین برای ایستگاه جلفا در هر شرایطی حداکثر بارش روزانه از رابطه خطی (۵) محاسبه می گردد.

GDP= 0.2893 * MONTHLY PRECIPITATION + 2.2633 (۵)

در روابط فوق GDP حداکثر بارش روزانه در ماه، AV TEMP میانگین دما، QFE PRES فشار سطح ایستگاه، AV VAP PRES فشار بخار میانگین،AV WIND میانگین سرعت باد و MONTHLY PRECIPITATION میزان بارش ماهانه می‌باشد. با توجه به این نتایج مشاهده می‌شود که روش‌های برنامه‌ریزی ژنتیک، مدل درختی M5 و شبکه عصبی مصنوعی با دقت بالا و مطلوبی قادر به پیش‌بینی حداکثر بارش روزانه در ایستگاه هواشناسی اهر و جلفا می‌باشند.

بحث و نتیجه‌گیری

در این تحقیق از داده‌های ایستگاه هواشناسی اهر و جلفا به عنوان ورودی جهت پیش‌بینی حداکثر بارش روزانه در ماه با استفاده از سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 استفاده گردید. با توجه به مقدار بالای ضریب همبستگی حاصل از سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 نتیجه می‌شود که هر سه روش مذکور می‌تواند با دقت بالایی میزان حداکثر بارش روزانه در ماه را تخمین بزنند. همچنین مدل درختی M5 به لحاظ سادگی، قابل فهم بودن و ارائه روابط خطی ساده در محدوده خاصی از داده‌های ورودی که بصورت روابط خطی ارائه می‌گردد، می تواند به عنوان یک روش جایگزین قابل رقابت با سایر روش‌ها از جمله شبکه عصبی مصنوعی مطرح شود. با در نظر گرفتن مطالعات سایر محققین در زمینه پیش‌بینی بارش با استفاده از روش شبکه عصبی مصنوعی مشاهده می‌شود که تحقیق حاضر از دقت مطلوبی نسبت به سایر تحقیقات برخوردار است. به عنوان مثال ترافالیس و همکاران (۲۰۰۵) بارش را با استفاده از روش شبکه عصبی مصنوعی با ($R^2 = ۰,۲$) تخمین زدند و یا فلاح قالهری و همکاران (۱۳۸۸) مقدار ($R^2 = ۰,۳۸$) را برای روش شبکه عصبی مصنوعی در پیش‌بینی بارش به دست آوردند. در حالی که در تحقیق حاضر مقدار ($R^2 = ۰,۷۴$) برای روش شبکه عصبی مصنوعی به دست آمد. پایین بودن ضریب همبستگی در دیگر مطالعات می تواند بدلیل وجود اشکال و ایراد در داده های ثبت شده تاریخی و یا ساختار و معماری شبکه عصبی مصنوعی تعریف شده باشد. نتایج برخی تحقیقات نیز تقریبا نزدیک به نتایج تحقیق حاضر می باشد. از جمله قلی زاده و دارند (۱۳۸۸) میزان بارش در تهران را با ($R = ۰,۸۸$) بدون اعمال الگوریتم ژنتیک در مدل شبکه‌های عصبی مصنوعی پیش‌بینی کردند و (فلاحی و همکاران، ۱۳۹۰) با به کارگیری مدل رگرسیون درختی، میزان بارش را در سه ایستگاه هواشناسی بندرانزلی، اراک و قم با ($R = ۰,۸۶$) تخمین زدند. با توجه به اینکه مدل درختی M5 یک الگوریتم پیش‌بینی عددی بوده و گره‌های درخت با خواص بیشینه خطاهای مورد انتظاری که به عنوان تابعی از انحراف استاندارد پارامترهای خروجی می باشد انتخاب می شوند، روابط خطی ارائه شده در برگ ها می توانند با دقت نسبتا زیادی مقادیر ویژگی هدف (متغییر وابسته هدف) را مدل نمایند با توجه به ساده و کارآمد بودن روش M5 که دقت قابل قبولی نیز در این مطالعه از خود نشان داد می‌توان کاربرد این روش را در مباحث مرتبط توصیه نمود.

منابع

۱. حلبیان، ا.ح. ۱۳۸۸. پیش آگاهی و برآورد بارش یزد با استفاده از شبکه های عصبی مصنوعی. تحقیقات کاربردی علوم جغرافیایی، جلد ۱۱، شماره ۱۴، ص ۲۸-۷.

۲. خلیلی، ن.، س.ر. خداشناس، ک. داوری و م. موسوی بایگی. ۱۳۸۷. پیش بینی بارش با استفاده از شبکه های عصبی مصنوعی. مجله علوم و صنایع کشاورزی، ویژه آب و خاک، جلد ۲، شماره ۱، ص۹۸-۸۹.

۳. سلطانی، ع.، م.ع. قربانی، ا.، فاخری فرد، ص. دربندی و د. فرسادی زاده. ۱۳۸۹. برنامه ریزی ژنتیک و کاربرد آن در مدل سازی فرآیند بارش-رواناب. مجله دانش آب و خاک، شماره ۴، ص ۷۱-۶۱.

۴. شریف زاک، م. و غ.ع. بارانی. ۱۳۹۰. پیش بینی دبی جریان در کانال های مرکب با استفاده از الگوریتم شبکه های عصبی. اولین کنفرانس ملی عمران و توسعه زیباکنار.

۵. علیخانزاده، امیر، ۱۳۸۵، داده کاوی، چاپ اول، نشر علوم رایانه، بابل.

۶. فاتحی مرج، ا. و م.ح. مهدیان. ۱۳۸۸. پیش بینی بارش پاییزه با استفاده از شاخص های انسو به روش شبکه عصبی در حوضه دریاچه ارومیه. پژوهش های آبخیزداری، شماره ۸۴، ص ۵۲-۴۲.

۷. فلاح قالهری، غ.ع. و ج. خوشحال. ۱۳۸۸. پیش بینی بارش بهاره استان خراسان رضوی بر اساس سیگنال های بزرگ مقیاس اقلیمی با استفاده از شبکه های عصبی مصنوعی. پژوهش های جغرافیای طبیعی، شماره ۶۹، ص ۱۳۳-۱۱۵.

۸. فلاح قالهری، غ.ع.، م. موسوی بایگی و م. حبیبی نوخندان. ۱۳۸۸. مقایسه نتایج به دست آمده از کاربرد سیستم استنباط فازی ممدانی و شبکه های عصبی مصنوعی در پیش بینی بارش فصلی، مطالعه موردی: منطقه خرسان. تحقیقات منابع آب ایران، سال پنجم، شماره ۲، ص ۵۲-۴۰.

۹. فلاحی، م.ر.، ه. ورواتی و س. گلیان. ۱۳۹۰. پیش بینی بارش با استفاده از مدل رگرسیون درختی به منظور کنترل سیل. پنجمین کنفرانس سراسری آبخیزداری و مدیریت منابع آب وخاک کرمان.

۱۰. قلی زاده، م.ح. و م. دارند. ۱۳۸۸. پیش بینی بارش ماهانه با استفاده از شبکه های عصبی مصنوعی (مورد: تهران). پژوهش های جغرافیای طبیعی، شماره ۷۱، ص ۶۳-۵۱.

11. Alberg, D., M. Last and A. Kindle. 2012. Knowledge discovery in data streams with regression tree methods. WIREs Data Mining Knowl Discov (2): 69-78.

12. Bhattacharya, B., Price, R. K., Solomatine, D. P., 2007. Machine Learning Approach to Modeling Sediment Transport. Journal of Hydraulic Engineering 133(4), 440-450.

13. Bhattacharya, B., Solomatine, D.P., 2005. Neural networks and M5 model trees in modeling water level–discharge relationship. Neurocomputing 63, 381–396.

14. Daga, M., Deo, M.C., 2009. Alternative data-driven methods to estimate wind from waves by inverse modeling, Nat Hazards 49, 293-310.

15. Ditthakit, P. and CH. Chinnarasri. 2012. Estimation of pan coefficient using M5 model tree. American Journal of Environmental Sciences 8 (2): 95-103.

16. Etemad-Shahidi, A., Bonakdar, L., 2009. Design of Rubble-Mound Breakwaters using M5′ Machine Learning Method. Submitted to Applied Ocean Research

17. Etemad-Shahidi, A., Mahjoobi, J., 2009. Comparison between M5′ Model Tree and Neural Networks for Prediction of Significant Wave Height. Submitted to Ocean Engineering.

18. Fallah Galhary, G.A., M. Habibi Nokhandan and M. Mousavi Baygi. 2010. Spring rainfall prediction based on remote linkage controlling using adaptive neuro-fazzy inference system (ANFIS). Theor Appl-Climato (101): 217-233.

19. Hakurta, P.G. 2008. Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. Meteorology and Atmospheric Physic (101): 93-108.

20. Londhe, S. N and Dixit, P. R., 2011. Forecasting Stream Flow Using Model Trees. International Journal of Earth Sciences and Engineering. 4(6): 282-285.

21. Mahjoobi, J., Etemad-Shahidi, A, 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. Applied Ocean Research. 30,172-177

22. Pal, M. 2006. M5 model tree for land cover classification. International Journal of Remote Sensing. 27(4), 20: 825-831.

23. Pal, M. and S. Deswal. 2009. M5 model tree based modeling of reference evapotranspiration. Hydrol. Process. (23): 1437-1443.
24. Quinlan, J.R. 1992. Learning with continuous classes. In proceedings AI,92 (Adams & Sterling, Eds), 343-348, Singapore: World Scientific
25. Sakhare, S., Deo, M.C., 2009. Derivation of wave spectrum using data driven methods, Marine Structures, 22, 594-609.
26. Satishkumar, S.K. and M. Rajib. 2012. Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation pattrns using Genetic Programming. Journal of Hydrology 454-455: 26-41.
27. Singh, K.K., M. Pal and V.P. Singh. 2010. Estimation of mean annual flood in Indian catchment using backpropagation neural network and M5 model tree. Water ressour manage 24 DOI: 10.1007/s11269-009-9535-x
28. Trafalis, T.B., B. Santosa and M.B. Richman. 2005. Learning networks in rainfall estimation. CMS (2): 229-251.

Monthly rainfall prediction using Artificial Neural Networks and M5 model tree (Case study: Stations of Ahar and Jolfa)

M. Taghi Sattari[1] and Farnaz Nahrein[2]

Abstract

Rainfall has been one of the most important agents in water cycle which has an effective rule in each region characters measurement. Prediction of month scale rainfall is important for main goals as torrent estimating, drought, run-off, sediment, irrigation programming and also manage the drainage basins. Rainfall measure prediction in each area mediated by punctual data measured of humidity, temperature, barograph manometers, wind speed and etc. the limitations such as unavailable enough data about rainfall measure on a scale of time and location and also complicated boundaries among meteorology agents related to rainfall ,caused to inexact and non trustable amount based on unusual manners. In this research ,firstly the description of different concepts of meteorology parameters on month scale in Ahar and Jolfa regions, EAST AZARBAIJAN, have been explained in which the entrance Artificial Neural Networks, Genetic Programming and M5 tree model have been defined too. Then, the best concept has been chosen for each model according to both R and RMSE statistics. In Ahar station Genetic Programming approach with (R=0.88) and (RMSE=3.32), also in Jolfa station Genetic Programming approach with (R=0.87) and (RMSE=3.79) presented the best results. The conclusion determined that each mentioned approaches presents the comparatively exact result for rainfall prediction in region but due to having simple liner models and understandable with M5 tree model, this approach would be considerate as an efficient application and substitutes for rainfall measurement.

Key words: Prediction of month scale rainfall, Artificial Neural Networks, Genetic Programing, M5 tree model, Ahar and Jolfa.

1- Water Engineering Department, Agriculture Faculty, University of Tabriz, Email: mtsattar@gmail.com
2- Msc Student, Water Engineering Department, Agriculture Faculty, University of Tabriz