

بهبود شبکه عصبی روش گروهی مدل‌سازی داده با استفاده از الگوریتم

حشره آب‌سوار در پیش‌بینی دما

شراره ملبوسی^{*}، مهدی خزاعی پور^۲، سمیرا شهرکی^۳

۱- دانشجوی دکتری، دانشگاه آزاد اسلامی واحد فردوس، فردوس

۲- استادیار، دانشگاه آزاد اسلامی واحد بیرجند، گروه کامپیوتر، بیرجند

۳- دانشجوی دکتری، دانشگاه آزاد اسلامی واحد بیرجند، بیرجند

چکیده

تغییرات اقلیمی آشکار پیش‌رو، تبدیل به دغدغه‌ای جدی برای جامعه بشری شده است و در این بین گرمایش جهانی یکی از گسترده‌ترین و مهم‌ترین مخاطرات زیست‌محیطی است بنابراین پیش‌بینی مناسب دما، اهمیت قابل توجهی در راستای سازگاری با تغییر اقلیم و کاهش آسیب‌پذیری‌ها در مقیاس‌های محلی دارد. برای این منظور در این تحقیق پیش‌بینی دما با یکی از روش‌های شبکه عصبی را پیشنهاد دادیم. پس از جمع‌آوری داده‌های ایستگاه اقدسیه در مرحله پیش‌پردازش پس از پاک‌سازی و نرمال‌سازی داده‌ها، عمل انتخاب ویژگی با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی انجام می‌شود، سپس در مرحله پس‌پردازش شبکه عصبی روش گروهی مدل‌سازی داده با استفاده از الگوریتم حشره آب‌سوار بهبود داده می‌شود تا پیش‌بینی دما به صورت بهینه انجام شود. نوآوری این تحقیق استفاده از الگوریتم حشره آب‌سوار در بهبود شبکه عصبی روش گروهی مدل‌سازی داده است. پارامترهایی در روش گروهی مدل‌سازی داده‌ها به عنوان متغیرهای تصمیم‌گیری تعریف می‌شوند در این تحقیق مقادیر بهینه این پارامترها، توسط الگوریتم حشره آب‌سوار تعیین شده تا پیش‌بینی دما با دقت بالایی انجام شود و در جهت مقایسه روش پیشنهادی، از شبکه عصبی پرسپترون چندلایه آموزش‌یافته با الگوریتم حشره آب‌سوار استفاده شده است. نتایج حاکی از میانگین مربعات خطای ۰۰۴۶۹ در روش پیشنهادی دارد.

کلید واژه‌ها: پیش‌بینی دما، شبکه عصبی روش گروهی مدل‌سازی داده، الگوریتم حشره آب‌سوار.

مقدمه

رديابی و پیش‌بینی نوسانات دما نقشی حیاتی در مطالعه الگوهای آب و هوایی آینده دارد. این روزها، تغییرات اقلیمی و گرم شدن کره زمین یکی از موضوعات داغ در سراسر جهان است زیرا تأثیر منفی آنها زندگی انسان‌ها را تغییر می‌دهد. پیش‌بینی می‌شود که تغییرات آب و هوایی خطر بلایای مرتبط با آب مانند سیل‌های شهری و خشکسالی‌های شدید را افزایش دهد یا تأثیر شدیدی بر کیفیت آب رودخانه‌ها در بسیاری از مناطق در سراسر جهان دارد. داشتن درک قوی از آینده، به ویژه در مورد تغییرات دما، در کمک به تصمیم‌گیرندگان برای ارزیابی و کاهش اثرات تغییرات آب و هوا و افزایش قابلیت اطمینان و پایداری زیرساخت‌ها بسیار مهم است. پیش‌بینی آب‌وهوا همواره به عنوان یک چالش بزرگ برای کارشناسان هواشناسی برای پیش‌بینی وضعیت جوی در آینده و شرایط آب و هوایی بوده است. پیش‌بینی دقیق آب‌وهوا در حوزه‌های مختلف مانند کشاورزی، آبیاری، تجارت دریایی و پیش‌بینی شرایط جوی می‌تواند زندگی منجر به حفظ جان انسان‌ها نسبت به بروز حوادث غیر مترقبه شود. از این‌رو در این پژوهش تلاش می‌شود تا با بهره‌گیری از شبکه عصبی روش گروهی مدل سازی داده و بهبود آن با الگوریتم حشره آب‌سوار به تخمین و پیش‌بینی دمای هوا در یک منطقه نمونه بپردازیم. ادامه ساختار این مقاله به این صورت است که در بخش دوم مروری بر پیشینه تحقیقات انجام شده خواهیم داشت و در بخش سوم شبکه عصبی روش گروهی مدل‌سازی داده، شبکه عصبی پرسپترون چند لایه، حشره آب‌سوار و تحلیل مؤلفه‌های اصلی بیان می‌شوند. در بخش چهارم مدل پیشنهادی به صورت پیش‌بینی دما به صورت گام به گام مطرح می‌شود و در بخش پنجم نتایج استفاده از شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار به عنوان روش پیشنهادی و شبکه عصبی پرسپترون چند لایه آموزش یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما بیان می‌شود و در نهایت در بخش آخر نتیجه‌گیری مطرح می‌گردد.

مروری بر پیشینه تحقیق

گوپتا و همکاران (۲۰۲۲) در مقاله خود به یک مدل مبتنی بر رگرسیون خطی چندگانه برای پیش‌بینی میانگین دمای یک روز پرداختند. آب و هوا پدیده‌ای است که هر روز بر همه

چیز و همه اطرافیان ما تأثیر می‌گذارد. پیش‌بینی آب و هوا برای دهه‌ها موضوع مهمی بوده است، زیرا محققان تلاش کرده‌اند تا با استفاده از تکنیک‌های هواشناسی سنتی، آب و هوا و تغییرات اقلیمی را پیش‌بینی کنند. با ظهور فناوری‌های مدرن و قدرت محاسباتی، می‌توان این کار را با کمک تکنیک‌های یادگیری ماشین انجام داد. هدف این مقاله پیش‌بینی آب و هوای یک منطقه با استفاده از داده‌ها و ویژگی‌های هواشناسی گذشته با استفاده از مدل رگرسیون خطی چندگانه است. عملکرد مدل ارزیابی شده و نتیجه‌گیری می‌شود. این مدل با موفقیت می‌تواند میانگین دمای یک روز را با خطای ۲.۸ درجه سانتیگراد پیش‌بینی کند. فونا و همکاران (۲۰۲۱) در مقاله خود به تخمین بارش باران بر اساس سنسجش از دور با استفاده از داده‌های حجیم به کمک چارچوب مسیریابی یکپارچه پرداختند. در این مقاله، جدیدترین تکنیک پیشرفته سنسجش از دور با مدل شبکه عصبی مصنوعی و کنترل منطق فازی مرتبط با مدل داده‌های حجیم با داده‌های موجود اداره هواشناسی هند برای سال آینده ایجاد شده است. نتایج نشان‌دهنده ارزیابی مناسب و کاهش خطا برای پیش‌بینی بارش در مدل پیشنهادی BIRSM و شبکه عصبی مصنوعی می‌باشد. چنگ و همکاران (۲۰۲۱) در مقاله خود به بررسی کاهش ویژگی بر اساس الگوریتم ژنتیک به منظور توسعه همزمان داده‌های هواشناسی در اینترنت اشیا صنعتی پرداختند. در این مقاله، جمعیت تکاملی به دو زیر جمعیت تقسیم شد: یکی از زیر جمعیت‌ها از افراد نخبه برای کمک به عملیات متقاطع برای افزایش سرعت همگرایی الگوریتم استفاده می‌کند و زیر جمعیت دیگر با معرفی یک جمعیت تصادفی، تنوع جمعیت را در روند تکامل متعادل می‌کند. سپس این دو زیر جمعیت با هم عملیات تکاملی را به پایان می‌رسانند. با استفاده از دو الگوریتم TSDFSO-AR و الگوریتم ARAGA، عملیات کاهش ویژگی بارش در داده‌های هواشناسی انجام شده است. نتایج نشان می‌دهد که الگوریتم پیشنهادی تنوع جمعیت را در طول تکامل حفظ کرده، عملکرد کاهش را بهبود بخشیده و سیستم اطلاعات را ساده نموده است. کاروان و سویکنز (۲۰۲۰) در مقاله خود با استفاده از الگوریتم LSTM مبتنی بر پیش‌بینی سری زمانی به ارائه روشی به منظور پیش‌بینی هوا پرداختند. در این مطالعه، یک تابع هزینه درجه دوم برای مسأله رگرسیون در نظر گرفته شده

است. بومی‌سازی تابع هدف با در نظر گرفتن یک تابع هزینه درجه دوم وزن‌دار انجام می‌شود که در آن نقطه نمونه‌های منطقه مورد آزمایش وزن بیشتری دارند. در این مقاله از دو طرح توزین براساس شباهت کسینوس بین نمونه‌های آموزش و آزمایش استفاده شده است. به منظور ارزیابی عملکرد روش پیشنهادی در شرایط مختلف آب و هوایی، آزمایشات در دو دوره زمانی مختلف از یک سال انجام می‌شود. نتایج نشان می‌دهد که T-LSTM منجر به عملکرد بهتر در کار پیش‌بینی تغییرات جوی می‌شود. دامودهاراوادهانی و همکاران (۲۰۱۹) در مقاله خود به پیش‌بینی بارش منطقه با استفاده از تکنیک‌های صافی‌نمایی مبتنی بر MapReduce پرداختند. در این مقاله، روش‌های هموارسازی نمایی ساده، Holt's Linear و هموارسازی نمایی Holt-Winter's با مدل محاسباتی MapReduce برای پیش‌بینی بارش منطقه‌ای پیشنهاد شده است. مطالعه تجربی بر روی دو مجموعه داده مختلف انجام شده است. اولین مورد مجموعه داده‌های باران هند است که شامل سال، ایالت و میزان بارش ماهانه به میلی‌متر است. دوم مجموعه داده‌های بارندگی ایالت تامیل نادو است که شامل سال، نواحی و میزان بارش ماهانه به میلی‌متر است. برای تأیید اعتبار این روش‌ها، اندازه‌گیری خطای MSE محاسبه شده است. نتایج نشان می‌دهد که روش صافی‌نمایی Holt-Winter's دقت بهتری را برای پیش‌بینی بارش دارد. سوارز و همکاران (۲۰۱۸) در مقاله خود به پیش‌بینی سری زمانی هوا با بهره‌گیری از ابرهای در حال تکامل و مدل‌های فازی پرداختند. در این مقاله تنوعی از روش هوشمند مبتنی بر ابر داده‌ای که به عنوان روش مبتنی بر ویژگی برای تجزیه و تحلیل داده استفاده شده است. هدف این مقاله این است که مدل‌های غیرخطی و متغیر زمان-داده محور را برای پیش‌بینی میانگین دمای ماهانه محاسبه کند. TEDA یک الگوریتم افزایشی است که تراکم داده‌ها و پراکندگی ابرها را در فضای داده در نظر می‌گیرد. این روش به دانش پیش‌بینی مجموعه داده و پارامترهای تعریف شده توسط کاربر نیاز ندارد. با این حال، اگر اطلاعاتی در مورد تعداد ابرها و قوانین موجود باشد، می‌توان آن را از طریق یک پارامتر بیان کرد. مقادیر گذشته حداقل، حداکثر و میانگین دمای ماهانه و همچنین مقادیر قبلی متغیرهای برون‌زا مانند ابر، بارندگی و رطوبت در تجزیه و تحلیل در نظر گرفته شده است. یک روش غیر پارامتری مبتنی

بر همبستگی اسپیرمن برای رتبه‌بندی و انتخاب مناسب‌ترین ویژگی‌ها و تأخیرهای زمانی برای پیش‌بینی دقیق‌تر ارائه شده است. مجموعه داده‌های این مقاله از ایستگاه‌های هواشناسی واقع در شهرهای اصلی برزیل مانند ساوپائولو، ماناوس، پورتو آلگره و ناتال به دست آمده است. این شهرها دارای ویژگی‌های خاص آب و هوایی شناخته شده‌اند. نتایج TEDA با نتایج ارائه شده توسط روش‌های تکامل‌یافته (Takagi-Sugeno eTS) و گسترش یافته (Takagi-Sugeno xTS) مقایسه شده‌اند. علاوه بر این، مجموعه‌ای از مدل‌های ابری و فازی و اپراتورهای تجمیع فازی برای ارائه پیش‌بینی‌های تک مقداره و دانه‌ای از سری زمانی ایجاد شده است. پیش‌بینی‌های دانه‌ای طیف وسیعی از مقادیر دما را منتقل می‌کند و ایده‌ای در مورد خطا و عدم اطمینان مرتبط با داده‌ها می‌دهد. راسل و همکاران (۲۰۱۷) در مقاله خود به بررسی کاربرد داده‌کاوی و یادگیری ماشین برای پیش‌بینی آب و هوا پرداختند. در این مقاله عملکرد پیش‌بینی داده‌ها و تکنیک‌های یادگیری ماشین با استفاده از رگرسیون بردار پشتیبانی و شبکه‌های عصبی مصنوعی به منظور پیش‌بینی آب و هوا به نمایش گذاشته شده است. برای انجام آزمایشات، مجموعه داده‌های آب و هوایی ۶ سال اخیر بارندگی و دمای منطقه کلان شهر چیتاگونگ از اداره هواشناسی بنگلادش جمع‌آوری شده است. یافته‌های این مطالعه نشان می‌دهد که الگوریتم رگرسیون بردار پشتیبانی نسبت به الگوریتم شبکه عصبی مصنوعی عملکرد بهتری در راستای پیش‌بینی بارندگی دارد. ساهو و همکاران (۲۰۱۷) در مقاله خود به ارائه یک رویکرد پیش‌بینی موازی با استفاده از تکنیک خوشه‌بندی K-means افزایشی پرداختند. در این مقاله یک روش تجزیه و تحلیل داده‌های آب و هوا با توجه به ماهیت داده‌های حجیم مربوط به داده‌های آب و هوا ارائه شده است. همچنین شرایط جوی به ترتیب تاریخ جمع‌آوری شده است. خوشه‌بندی سنتی k-means برای تشکیل خوشه‌هایی استفاده می‌شود که نشان دهنده ارتباط بین تاریخ‌های مربوط به داده‌های آب و هوایی سال جاری و سال قبل است. چنین خوشه‌هایی بر اساس داده‌های قبلی، شرایط جوی شرایط آب و هوایی یک ساله را پیش‌بینی می‌کنند. الگوریتم خوشه‌بندی k-means افزایشی برای پردازش پارامترهای آب و هوای سال جاری به عنوان داده‌های جدید استفاده می‌شود و نشان می‌دهد که شرایط آب و هوایی محاسبه شده برای نمایش شرایط جوی

حشره‌های آب‌سوار برای انتقال اطلاعات مختلف از ارتباطات موج‌دار استفاده می‌کنند. این سیستم برای شناسایی شکارچیان و طعمه‌های احتمالی استفاده می‌شود. آنها امواج را با دامنه‌ها، طول و فرکانس‌های مختلف توسط نوسان پاها روی سطح آب تولید می‌نمایند. هر یک از این سیگنال‌ها مانند اظهار عشق، دفع کردن، تعیین موقعیت طعمه و غیره برای اهداف خاصی بکار برده می‌شوند. آنها امواج‌های ارتباطی را از طریق گیرنده‌های حسی روی پاهایشان دریافت می‌کنند و به وضعیت واکنش نشان می‌دهند. برخلاف ماده‌های بی‌میل، نرها بسیار مشتاق به جفت شدن هستند. نرها سیگنال‌های فراخواندن را ارسال می‌کنند و ماده‌ها به آنها به صورت سیگنال‌های جذب یا دفع پاسخ می‌دهند. اگر یکدیگر را جذب کنند، آنها جفت خواهند شد. در غیر این صورت، نر متصل به اجبار می‌شود. هنگامی که حشرات درون آب می‌افتند و برای زنده ماندن تلاش می‌کنند، حشرات آب‌سوار سیگنال‌های موج را دریافت می‌کنند و آن را شکار می‌کنند. پنج گام اصلی تولد، تولید قلمرو، جفت‌شدن، تغذیه و مرگ را به طور ریاضی مدل می‌نماید.

گام اول تولد اولیه: در تولد، حشرات آب‌سوار با تخم‌های توزیع شده در دریاچه متولد می‌شوند. در اینجا، توزیع تصادفی به صورت رابطه (۱) فرض می‌شود.

$$WS_i^0 = Ub + (Ub - Lb)rand \quad i = 1, 2, \dots, nws \quad (1)$$

که WS_i^0 موقعیت اولیه حشره آب‌سوار i ام است. Ub و Lb به ترتیب کران‌های بالا و پایین مناظر با مقادیر ماکزیمم و مینیمم مجاز را نشان می‌دهند. $rand$ عدد تصادفی در بازه $[0, 1]$ تولید می‌نماید، nws تعداد حشرات آب‌سوار است. حشرات آب‌سوار اولیه با استفاده از تابع هدف ارزیابی می‌شوند تا مقدار برازندگی موقعیت‌شان روی دریاچه به دست آید.

گام دوم ساخت قلمرو: حشرات آب‌سوار برای زنده ماندن، جفت‌شدن و تغذیه، قلمروها را حفظ می‌کنند. برای ساخت تعداد nt قلمرو، تکنیک زیر برای انتساب حشرات آب‌سوار به قلمروها بکار برده می‌شود. در ابتدا، حشرات آب‌سوار براساس برازندگی‌شان مرتب می‌شوند و درون $\frac{nws}{nt}$ گروه مرتب تقسیم می‌شوند. عضو j ام هر گروه به قلمرو j ام تخصیص داده می‌شود که $j=1, 2, \dots, nt$. از این‌رو، درون هر قلمرو تعداد $\frac{nws}{nt}$ حشره آب‌سوار زندگی می‌نماید. از آنجایی که ماده‌ها معمولاً

مشابه تحت یکی از خوشه‌های موجود قرار می‌گیرد. اونال و همکاران (۲۰۱۷) در مقاله خود به تجزیه و تحلیل داده‌های جوئی و تشخیص خطای حسگر با استفاده از یک چارچوب مبتنی بر اینترنت اشیا، داده‌های حجیم و یادگیری ماشین پرداختند. در این مطالعه، خوشه‌بندی هوا و تشخیص ناهنجاری حسگر با استفاده از یک مجموعه داده در دسترس عمومی اجرا شده است. جزئیات پیاده‌سازی هر لایه چارچوب (دریافت، پردازش داده‌ها، آموزش و تصمیم‌گیری) برای منطقه مورد استفاده جهت تخمین شرایط آب و هوایی در نظر گرفته شده است. مدل یادگیری در داخل کتابخانه خوشه‌بندی مبتنی بر Scikit-Learn اجرا شده است. نتایج تجزیه و تحلیل داده‌ها نشان می‌دهد که استخراج اطلاعات معنی‌دار جوئی از یک مجموعه داده نسبتاً پیچیده با استفاده از چارچوب پیشنهادی امکان‌پذیر است.

مروری بر الگوریتم‌ها

در این بخش الگوریتم‌های حشره آب‌سوار، تحلیل مؤلفه‌های اصلی، شبکه‌های عصبی روش گروهی مدل‌سازی داده‌ها و پرسپترون چندلایه بررسی می‌شوند.

الگوریتم حشره آب‌سوار

الگوریتم حشره آب‌سوار در سال ۲۰۲۰ توسط کاوه و اسلام‌لو با الهام از زندگی حشره آب‌سوار ارائه شده است. آنها در بالای سطح آب زندگی می‌کنند در واقع، استفاده از تنش سطحی (جذب ذرات و آب در لایه سطحی) با پاهای آبرگیز باعث شده است تا در بالای سطح آب بمانند. آنها برای دفاع از دارایی‌های خود شامل مواد غذایی و منابع جفت شدن، قلمروهایی را ایجاد می‌کنند. منابع غذایی مورد نیاز ماده‌ها هستند که تاثیر مستقیمی روی تعداد تخم‌ها و زنده ماندنشان دارد. منابع جفت شدن مطلوب نرها می‌باشند که فاکتور کلیدی در تولیدمثل موفق‌شان است. در هر قلمرو معمولاً یک نر بالغ و چند ماده هستند. نر مقیم اصطلاحاً "سنگ کلیدی" نامیده می‌شود که فعالیت جفت‌شدن را انجام می‌دهد. در صورتی که حشره‌های آب‌سوار بیگانه وارد قلمروهایشان شوند، نرها رفتار قلمرویی خصمانه و خشونت‌آمیز را برای دفع مزاحمان نشان خواهند داد. حمله بین متجاوزین به قلمرو و ساکنین در قلمرو شدید است و ممکن است منجر به کشتن شود.

باشد، آن باید به سمت بهترین زیستگاه حاوی بالاترین برازندگی حرکت نماید. برای این هدف، رابطه (۱) برای انتقال به موقعیت جدید اطراف بهترین حشره دریاچه (WS_{BL}^t) شامل منابع غذایی خوب تعریف می‌شود.

$$WS_i^{t+1} = WS_i^t + 2rand.(WS_{BL}^t - WS_i^t) \quad (۴)$$

گام پنجم مرگ و موفقیت: به منظور تعیین نتیجه فرآیند دستیابی به غذا، تابع هدف ارزیابی می‌شود و با موقعیت قبلی مقایسه می‌شود. اگر برازندگی جدید کمتر است، حشره آب‌سوار کشته خواهد شد، زیرا آن نه تنها نمی‌تواند غذا را پیدا نماید بلکه همچنین خطر نبرد با حشرات آب‌سوار قلمرو مقصد را افزایش می‌دهد. در این حالت، نوزاد حشره بالغ جدید، موفق به کشتن حشره آب‌سوار سنگ کلیدی خواهد شد، همان‌طور که در رابطه (۵) نشان داده شده است.

$$WS_i^{t+1} = Lb_j^t + 2rand(Ub_j^t - Lb_j^t) \quad (۵)$$

که Ub_j^t و Lb_j^t به ترتیب مقادیر ماکزیمم و مینیمم موقعیت حشره‌های آب‌سوار درون قلمرو j ام هستند. به عبارت دیگر، آنها مرزهای قلمرو حشره آب‌سوار کشته شده را تعیین می‌نمایند. گام ششم شرط خاتمه الگوریتم حشره آب‌سوار: در گام آخر الگوریتم، اگر شرط خاتمه برقرار باشد، الگوریتم متوقف می‌شود و بهترین موقعیت تجربه شده را گزارش می‌نماید. اما اگر شرط خاتمه برقرار نباشد، آن به گام جفت‌گیری برای حلقه جدید چرخه زندگی و تولید قلمرو بازمی‌گردد.

تحلیل مؤلفه‌های اصلی

الگوریتم تحلیل مؤلفه‌های اصلی در سال ۱۹۹۱ معرفی شد. از این الگوریتم برای کاهش بُعد استفاده می‌شود تا بتواند زیرفضایی با بردارهای متعامد پیدا کند که در آن زیرفضا، پراکندگی داده‌ها را به بهترین حالت نشان دهد. در فضای جدید بردارها مستقل از یکدیگر عمل می‌کنند و هیچ‌گونه وابستگی بین آن‌ها وجود ندارد. پس از مشخص شدن بردارها، تمامی تصاویر به این زیرفضا منتقل می‌شوند.

روش گروهی مدل‌سازی داده

به طور کلی برای مدل‌کردن سیستم‌های پیچیده‌ای که شامل مجموعه‌ای از داده‌ها با چندین ورودی و یک خروجی

بهترین مکان هر قلمرو را برای غذایی پیدا می‌نمایند، در هر قلمرو موقعیت‌های با بهترین و بدترین برازندگی به ترتیب به عنوان موقعیت ماده (جستجوی زیستگاه بهینه) و نر (سنگ کلیدی) در نظر گرفته می‌شوند. در کل الگوریتم، فضای جستجو به صورت دریاچه‌ای که حاوی قلمروهای مختلف راه‌حل‌ها است، تعریف می‌شود و غذا به عنوان تابع هدف بکار برده می‌شود.

گام سوم جفت‌شدن: جفت‌شدن فرآیند مهمی در زندگی حشرات آب‌سوار است. نرهای سنگ کلیدی امواج فراخوانی را ارسال می‌کنند و ماده با ارسال سیگنال‌هایی موج جذب یا دفع پاسخ می‌دهد. احتمال ارسال پاسخ جذب برابر p در نظر گرفته می‌شود و بنابراین، احتمال پاسخ دفع برابر $1-p$ می‌باشد. از آنجایی که پاسخ ماده‌ها به سادگی تعیین نمی‌شود، $p=50\%$ در نظر گرفته می‌شود. اگر ماده سیگنال جذب بفرستد، آنها به سمت یکدیگر حرکت خواهند کرد و جفت خواهند شد. یک موج دایره را در نظر بگیرید. اگر ماده درخواست را رد نماید، نر را مجبور می‌کند و ماده از او دور می‌شود. حشره نر ممکن است جفت یا دفع شود یا دور شود، موقعیت جدید حشره نر با رابطه (۲) بهنگام خواهد شد.

که WS_i^t موقعیت حشره آب‌سوار i ام در تکرار t ام الگوریتم است. $Rand$ عدد تصادفی در بازه ۰ تا ۱ است. R برداری است که نقطه شروع در موقعیت نر (WS_i^{t-1}) و نقطه انتهایش در موقعیت ماده در قلمرو یکسان (WS_F^{t-1}) است. این ماده می‌تواند متناسب با برازندگی‌اش براساس مکانیزم‌های انتخاب مانند چرخ انتخاب رولت، انتخاب شود. طول R برابر فاصله اقلیدسی بین حشرات آب‌سوار نر (WS_i^{t-1}) و ماده (WS_F^{t-1}) (شعاع موج) است همان‌طور که در رابطه (۳) نشان داده شده است.

$$R = WS_F^{t-1} - WS_i^{t-1} \quad (۳)$$

گام چهارم تغذیه: همان‌طور که قبلاً بیان شد، چه فرآیند جفت‌گیری با موفقیت اجرا شود و چه نه، فرآیند مقدار زیادی انرژی را صرف می‌نماید. بنابراین، در موقعیت جدید، حشرات آب‌سوار برای منابع غذا جستجو می‌نمایند. همان‌طور که گفته شد، برای ارزیابی موقعیت برای موجود بودن غذا، آن تابع هدف را ارزیابی می‌نماید. اگر مقدار تابع هدف بزرگتر از حالت قبلی است، آن به معنای آن است که آن غذایی را برای بازیابی یافته است. اما اگر مقدار هدف کمتر از حالت قبلی

شبکه عصبی پرسپترون چندلایه

شبکه‌های عصبی به‌عنوان یک تکنولوژی کاربردی ظاهر شده‌اند که به صورت مؤفقت‌آمیزی در زمینه‌های متفاوت به‌کار برده می‌شوند. مهم‌ترین مزیت شبکه عصبی قابلیت خود تطبیقی، خودسازماندهی و عملیات بلادرنگ و غیره است. ساختار این شبکه شامل یک لایه ورودی، یک لایه میانی و یک لایه خروجی است. در هر لایه، یک یا چند عنصر پردازش‌گر (نورون)، وجود دارد که با تمامی نورون‌های لایه بعدی با اتصالات وزن‌دار، به هم مربوط می‌شوند. تعداد نورون‌های لایه‌های ورودی و خروجی، بستگی به تعداد متغیرهای ورودی و خروجی مدل دارد؛ ولی انتخاب تعداد نورون‌های لایه میانی به صورت سعی و خطا تعیین می‌شود که در مسأله ما معمولاً بهترین جواب تشخیص هویت با تعداد نورون‌های ۱۰ در لایه پنهان به دست می‌آید.

در شبکه‌های عصبی، نورون‌های هر لایه، به کلیه نورون‌های لایه قبل، از طریق یک اتصال جهت‌دار، مرتبط می‌شوند. به هر یک از این اتصالات، وزنی داده می‌شود که مقدار آن، تعیین‌کننده تأثیر هر نورون روی نورون لایه خروجی است. مجموع وزنی مقادیر ورودی به هر نورون، محاسبه و در یک تابع ریاضی، قرار می‌گیرد و خروجی نورون از طریق این تابع، محاسبه می‌شود. این تابع ریاضی را اصطلاحاً تابع فعال‌سازی می‌گویند. توابع سیگموئیدی و خطی، رایج‌ترین توابع مورد استفاده برای شبکه‌های چندلایه پیش‌رونده هستند. در این تحقیق از تابع تانژانت هایپربولیک در نورون‌های لایه پنهان و نورون‌های لایه خروجی استفاده شده است. وزن‌های ارتباط دهنده نورون‌های شبکه، با آموزش تعیین می‌شوند. الگوریتم‌های بهینه‌سازی هوشمند یکی از تکنیک‌های بهینه‌سازی قدرتمند هستند که می‌توانند دستیابی به وزن‌های بهینه‌شده در شبکه عصبی را میسر سازند.

مراحل روش پیشنهادی

هدف این بخش ارائه روش تحقیق و بهبود شبکه عصبی روش گروهی مدل‌سازی داده‌ها توسط الگوریتم حشره آب سوار به عنوان مدل پیشنهادی است.

می‌باشند می‌توان از چندجمله‌ای کولموگوروف-گابور در رابطه (۶) استفاده نمود.

$$\sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \dots \quad (6)$$

که در آن $x=(x_1, x_2, \dots, x_n)$ بردارهای ورودی، y خروجی مدل و a_i ضرایب چندجمله‌ای می‌باشند. چندجمله‌ای‌های کولموگوروف-گابور با استفاده از چندجمله‌ای‌های درجه دوم تقریب زده می‌شوند. این چندجمله‌ای‌های درجه دوم بر اساس ترکیبات دوتایی ورودی‌های شبکه ساخته می‌شوند. الگوریتم GMDH با استفاده از این ایده به عنوان یادگیری برای مدل‌سازی سیستم‌های پیچیده معرفی شده است. شبکه عصبی GMDH دارای ساختار یک شبکه چندلایه و رو به جلو می‌باشد و در برگیرنده مجموعه‌ای از نورون‌ها است که از پیوند جفت‌های مختلف ورودی از طریق یک چندجمله‌ای درجه دوم به وجود می‌آیند. هر لایه در این شبکه از یک یا چند واحد پردازشگر تشکیل شده است که هر یک از آنها دارای دو ورودی و یک خروجی می‌باشند. این واحدها عملاً نقش اجزای سازنده مدل را ایفا می‌نمایند و به شکل چندجمله‌ای درجه دوم فرض می‌شوند (رابطه (۷)).

$$\hat{y}_n = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2 \quad (7)$$

پارامترهای مجهول الگوریتم GMDH ضرایب چندجمله‌ای رابطه (۶) می‌باشد. جهت محاسبه y برای هر بردار ورودی $x=(x_1, x_2, \dots, x_n)$ بر اساس رابطه (۷)، باید میانگین مربعات خطا (رابطه (۸)) به حداقل برسد.

$$e = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (8)$$

برای یافتن مقدار حداقل خطا، از مشتق جزئی رابطه (۷) استفاده می‌شود. با جایگذاری از رابطه (۶) در این مشتق جزئی معادله ماتریسی $(Aa=y)$ به دست می‌آید. در این معادله A مطابق رابطه (۹) می‌باشد.

$$A = \begin{bmatrix} 1 & x_{1p} & x_{2p} & x_{1p}^2 & x_{2p}^2 & x_{1p}x_{2p} \\ 1 & x_{2p} & x_{2q} & x_{2p}^2 & x_{2q}^2 & x_{2p}x_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{np} & x_{nq} & x_{np}^2 & x_{nq}^2 & x_{np}x_{nq} \end{bmatrix} \quad (9)$$

یک روش حل برای این معادله ماتریسی $(Aa=y)$ ، استفاده از روش تجزیه مقدار تکین است. در صورت استفاده از این روش مجهول a از رابطه $(Aa=y)$ محاسبه می‌گردد.

جمع‌آوری داده‌ها

پایگاه داده شامل داده‌های ایستگاه‌های هواشناسی اقدسیه است شامل ۲۷۲۱ نمونه با ۲۳ ویژگی است که مربوط به ۸ سال گذشته است که ۸۰٪ داده‌ها برای آموزش مدل‌ها و ۲۰٪ از داده‌ها برای آزمایش مدل‌ها استفاده شده است.

طوری که همبستگی بین ویژگی‌ها را از بین برده و ویژگی‌های با پراکندگی بالا انتخاب می‌شود. در این تحقیق با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی از میان ۲۳ ویژگی موجود در پایگاه داده تعداد ۶ ویژگی انتخاب شده‌اند.

مروری بر روش‌های ارزیابی خطا

هدف از ارزیابی، بررسی دقت تخمین مدل‌های ارائه شده با نتایج اندازه‌گیری شده است. برای ارزیابی مدل‌ها از روش میانگین مربع خطاها (MSE)، جذر میانگین مربع خطاها (RMSE)، میانگین قدرمطلق خطاها (MAE) و مجموع مربعات خطاها (SSE) می‌توان استفاده کرد که از روابط زیر به دست می‌آیند.

$$MSE = \sum_{i=1}^n \frac{(t_i - y_i)^2}{n} \quad (11)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(t_i - y_i)^2}{n}} \quad (12)$$

$$MAE = \sum_{i=1}^n \frac{|t_i - y_i|}{n} \quad (13)$$

$$SSE = \sum_{i=1}^n (t_i - y_i)^2 \quad (14)$$

پاک‌سازی داده‌ها

در بسیاری از کاربردهای دنیای واقعی کاوش داده‌ها، حتی با وجود مقدار داده‌های حجیم و فضای ذخیره‌سازی مناسب، ممکن است در نمونه‌های موجود، مقادیری از داده‌ها، از دست‌رفته (گم‌شده) باشند. اما مشکل از آنجا آغاز می‌شود که برای مجموعه داده‌های بزرگ نمی‌توان از مقادیر از دست‌رفته چشم‌پوشی کرد. یک راه‌حل برای جایگزینی و پاک‌سازی مقادیر از دست‌رفته با مقادیر ثابت است در این تحقیق از مقدار میانگین برای مقادیر از دست‌رفته در هر ویژگی استفاده شده است. به عبارت دیگر براساس مقادیر موجود برای هر ویژگی میانگین محاسبه شده و در نمونه‌های فاقد مقدار جایگزین می‌شود.

مراحل روش پیشنهادی

از چالش‌های اساسی در روش گروهی مدل‌سازی داده‌ها، تعیین حداکثر تعداد لایه‌ها، تعیین حداکثر تعداد نورون‌ها در هر لایه، پارامتر فشار انتخاب و نرخ داده‌های آموزشی است که توسط کاربر انجام می‌شود و ممکن است مقادیر بهینه‌ای توسط کاربر انتخاب نشود و دقت در پیش‌بینی کاهش یابد. بنابراین در این تحقیق پارامترهای ذکر شده در روش گروهی مدل‌سازی داده‌ها به عنوان متغیرهای تصمیم‌گیری تعریف می‌شوند و میانگین مربعات خطا به عنوان تابع هدف تعریف و توسط الگوریتم حشره آب‌سوار مقادیر بهینه پارامترها تعیین شده تا پیش‌بینی دما با دقت بالایی انجام شود. در این بخش ابتدا نحوه تعریف متغیرهای تصمیم‌گیری و تابع هدف برای بهبود شبکه عصبی روش گروهی مدل‌سازی داده‌ها بیان می‌شود و مراحل تعیین پارامترهای بهینه توسط الگوریتم حشره آب‌سوار مطرح می‌شود. در جدول (۱) متغیر تصمیم از چهار بعد حداکثر تعداد لایه‌ها، حداکثر تعداد نورون‌ها در هر لایه، پارامتر فشار انتخاب و نرخ داده‌های آموزشی تشکیل شده است.

نرمال‌سازی داده‌ها

باتوجه به یکسان نبودن بازه تغییرات ویژگی‌ها و همچنین واحدهای متفاوت متغیرها، مقادیر بزرگ‌تر تأثیر بیشتری بر توابع مورد استفاده دارند که لزوماً به معنی مهم‌تر بودن آنها نیست. برای رفع این مشکل نرمال‌سازی داده‌ها انجام می‌شود. در این تحقیق با نرمال‌سازی خطی، داده‌ها به بازه $[-1, 1]$ نرمال‌سازی شده‌اند.

$$X = 2 \times \frac{x - \min(x)}{\max(x) - \min(x)} - 1 \quad (10)$$

که در آن $\min(x)$ کمینه بردار ورودی x و $\max(x)$ بیشینه بردار ورودی x بوده و X نرمال شده آن است.

انتخاب ویژگی

تحلیل مؤلفه‌های اصلی یکی از انواع روش‌های تحلیل داده‌های چندمتغیره است که هدف اصلی آن تقلیل بُعد مسأله مورد مطالعه است. یکی از کاربردهای مهم تحلیل مؤلفه‌های اصلی در رگرسیون است. الگوریتم تحلیل مؤلفه‌های اصلی داده‌ها را از فضای ورودی به فضایی جدید نگاشت می‌کند به

جدول ۱- چهار بعد متغیر تصمیم

حداکثر تعداد نورون‌ها	حداکثر تعداد لایه‌ها	فشار انتخاب	نرخ داده‌های آموزشی
-----------------------	----------------------	-------------	---------------------

مذکور را به تابع هدف ارسال می‌کند در این وضعیت تابع هدف بر اساس مقادیر ارسالی شبکه عصبی روش گروهی مدل‌سازی داده‌ها را ایجاد و آموزش می‌دهد و پس از آموزش ورودی‌های آموزشی به شبکه عصبی اعمال شده تا خروجی تعیین شود. در این وضعیت مقدار میانگین مربعات خطا بر اساس خروجی‌های شبکه عصبی و خروجی‌های ایده‌آل محاسبه شده و به عنوان مقدار شایستگی حشره آب‌سوار در نظر گرفته می‌شود. در جدول (۲) برای حشره آب‌سوار مورد نظر حداکثر تعداد نورون‌ها برابر ۱۵، حداکثر تعداد لایه‌ها برابر ۸، فشار انتخاب برابر ۰.۶۵ و نرخ داده‌های آموزشی برابر ۰.۸ در نظر گرفته شده است.

مرحله تکرار: در این مرحله عملیات زیر تکرار می‌شوند: گام ساخت قلمرو: در این گام تعدادی قلمرو از حشرات آب‌سوار ایجاد می‌شود که در هر قلمرو تعداد یکسانی حشره آب‌سوار قرار می‌گیرد. گام جفت شدن: در این گام حشره نر امواج فراخوانی را ارسال می‌کند و حشره ماده آب‌سوار سیگنال موج جذب را ارسال می‌کند. جدول (۳) موقعیت حشره آب‌سوار نر را قبل و بعد از جفت شدن نشان می‌دهد. در این مرحله عملیات زیر در هر نسل از اجرای الگوریتم تکرار می‌شوند.

در رابطه (۱۵) تابع میانگین مربعات خطا به عنوان تابع هدف برای تعیین بهترین مقادیر پارامترهای مذکور در شبکه عصبی روش گروهی مدل‌سازی داده‌ها در الگوریتم حشره آب‌سوار در نظر گرفته شده است.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{targets}_i - \text{outputs}_i)^2 \quad (15)$$

$$= \frac{1}{n} \sum_{i=1}^n e_i^2$$

که در آن targets_i خروجی ایده‌آل و outputs_i خروجی شبکه عصبی روش گروهی مدل‌سازی داده‌ها است. در این بخش مراحل بهبود شبکه عصبی روش گروهی مدل‌سازی داده‌ها توسط الگوریتم حشره آب‌سوار بیان می‌شود. این الگوریتم دارای دو مرحله آماده‌سازی اولیه و مرحله تکرار است.

مرحله آماده‌سازی اولیه: در این مرحله جمعیتی از حشره‌های آب‌سوار به صورت تصادفی ایجاد می‌شوند. موقعیت هر حشره آب‌سوار شامل چهار مؤلفه حداکثر تعداد لایه‌ها، حداکثر تعداد نورون‌ها در هر لایه، پارامتر فشار انتخاب و نرخ داده‌های آموزشی می‌باشد که به صورت تصادفی مقداردهی شده‌اند. هر حشره آب‌سوار پس از مقداردهی پارامترهای

جدول ۲- موقعیت تصادفی حشره آب‌سوار در الگوریتم حشره آب‌سوار

حداکثر تعداد نورون‌ها	حداکثر تعداد لایه‌ها	فشار انتخاب	نرخ داده‌های آموزشی
۱۵	۸	۰.۶۵	۰.۸

جدول ۳- موقعیت جدید حشره آب‌سوار قبل و بعد از جفت شدن در الگوریتم حشره آب‌سوار

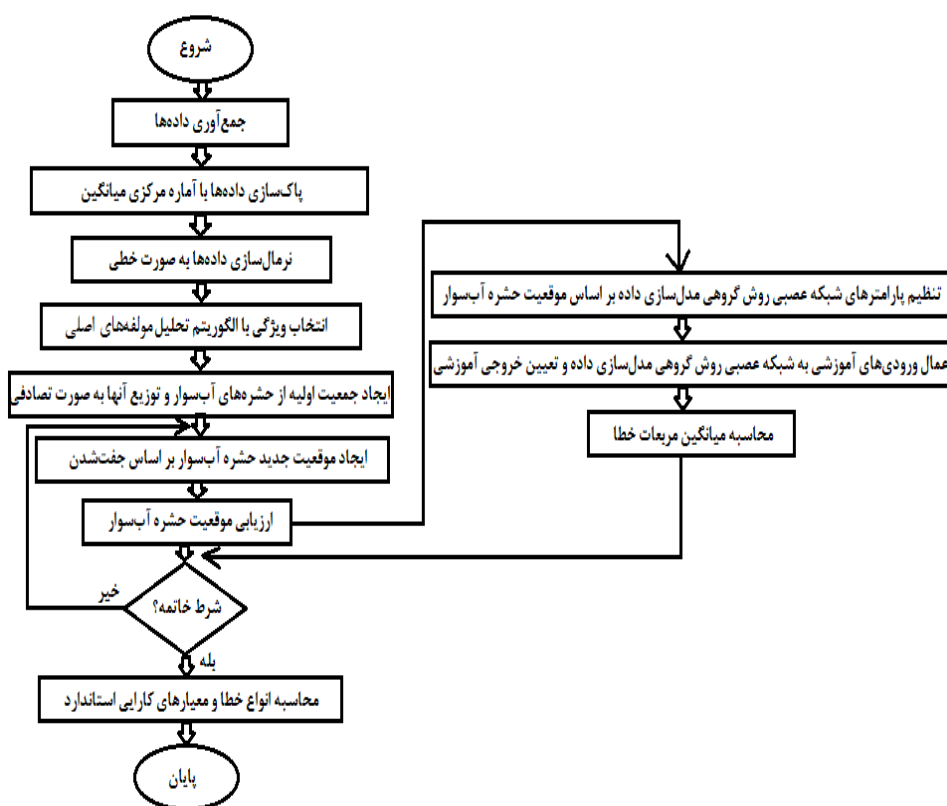
موقعیت حشره آب‌سوار قبل از جفت شدن:			
حداکثر تعداد نورون‌ها	حداکثر تعداد لایه‌ها	فشار انتخاب	نرخ داده‌های آموزشی
۱۵	۸	۰.۶۵	۰.۸

موقعیت حشره آب‌سوار بعد از جفت شدن:			
حداکثر تعداد نورون‌ها	حداکثر تعداد لایه‌ها	فشار انتخاب	نرخ داده‌های آموزشی
۱۰	۱۲	۰.۸۷	۰.۷۳

محاسبه شده و به عنوان مقدار شایستگی حشره آب‌سوار در نظر گرفته می‌شود.

مراحل فوق تا برآورده شدن شرایط خاتمه تکرار می‌شود. خروجی الگوریتم حشره آب‌سوار عضوی از جمعیت حشره آب‌سوار است که بهترین مقادیر چهار مؤلفه حداکثر تعداد لایه‌ها، حداکثر تعداد نوروها در هر لایه، پارامتر فشار انتخاب و نرخ داده‌های آموزشی را برای شبکه عصبی روش گروهی مدل‌سازی داده‌ها دارد. در شکل (۱) فلورچارت روش پیشنهادی نمایش داده شده است.

ارزیابی موقعیت جدید حشره آب‌سوار: پس از تعیین موقعیت جدید حشره آب‌سوار با عمل جفت‌شدن، پارامترهای مذکور را به تابع هدف ارسال می‌کند در این وضعیت تابع هدف بر اساس مقادیر ارسالی شبکه عصبی روش گروهی مدل‌سازی داده‌ها را ایجاد و آموزش می‌دهد و پس از آموزش ورودی‌های آموزشی به شبکه عصبی اعمال شده تا خروجی تعیین شود. در این وضعیت مقدار میانگین مربعات خطا براساس خروجی‌های شبکه عصبی و خروجی‌های ایده‌آل



شکل ۱- فلورچارت روش پیشنهادی

شبیه‌سازی در مرحله پیش‌پردازش (پاک‌سازی، نرمال‌سازی و انتخاب ویژگی) و پس‌پردازش (شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار و شبکه عصبی پرسپترون چندلایه) در محیط نرم افزار متلب ۲۰۲۱ انجام شده است.

نتایج شبیه‌سازی

این بخش به بررسی نتایج حاصل از مدل پیشنهادی و الگوریتم‌های دیگر در پیش‌بینی دما می‌پردازد لذا ابتدا نتایج مربوط به استفاده از شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار می‌پردازیم و در ادامه نتایج استفاده از شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب‌سوار بررسی می‌شود. کلیه مراحل

پارامترهای اندازه جمعیت حشره‌های آب‌سوار برابر ۵۰ و حداکثر تعداد دفعات تکرار برابر ۱۰۰۰ در نظر گرفته شده است و نتایج براساس انواع خطا مطابق جدول (۴) است.

نتایج پیش بینی دما با روش پیشنهادی

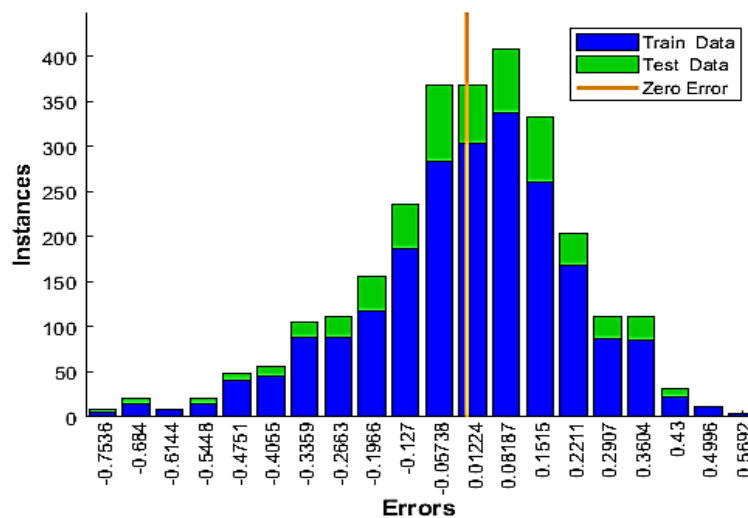
در استفاده از شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما،

جدول ۴- انواع خطا در روش پیشنهادی در پیش‌بینی دما

	MSE	RMSE	MAE	SSE
Train Data	۰.۰۴۷۱	۰.۲۱۷۱	۰.۱۶۷۳	۱۰۲.۵۶
Test Data	۰.۰۴۶۲	۰.۲۱۴۹	۰.۱۶۷۴	۲۵.۱۱۷
All Data	۰.۰۴۶۹	۰.۲۱۶۶	۰.۱۶۷۳	۱۲۷.۶۸۱

خروجی های شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما با مربع‌های آبی رنگ به ازای داده‌های آموزشی نمایش داده شده است. محور افقی اندیس داده‌ها و محور عمودی مقادیر دما هدف و پیش‌بینی شده نمایش داده شده است.

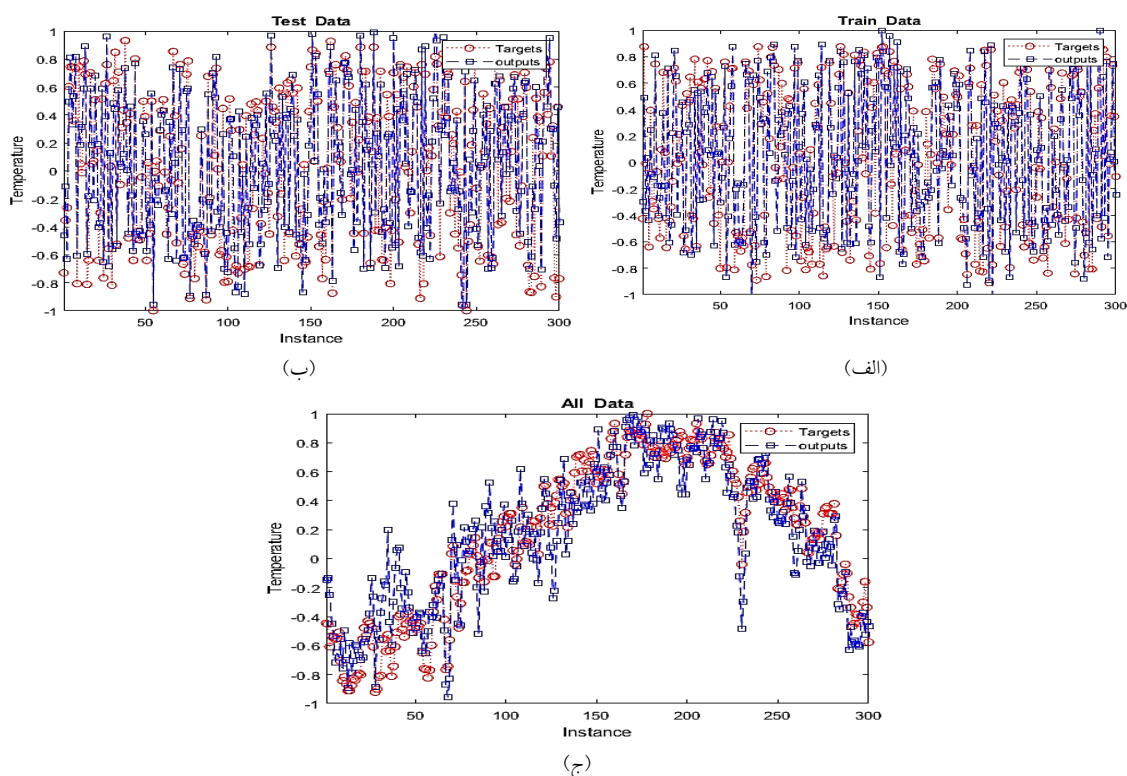
در شکل (۲) هیستوگرام خطا به ازای داده های آموزشی با رنگ آبی و به ازای داده های آزمایشی با رنگ سبز نمایش داده شده است محور افقی مقدار خطا و محور عمودی فراوانی خطا در شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار را نشان می دهد. در شکل (۳-الف) خروجی های هدف با دایره‌های قرمز و



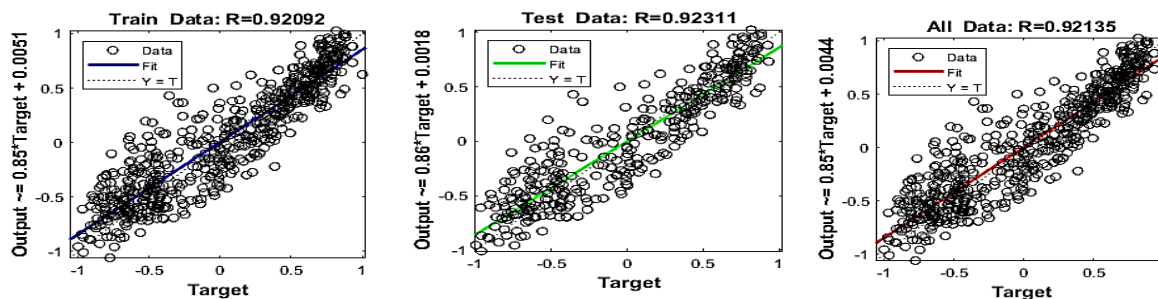
شکل ۲- هیستوگرام خطا به ازای داده‌های آموزشی و آزمایشی در GMDH بهبود یافته با الگوریتم حشره آب‌سوار

شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما با مربع‌های آبی رنگ به ازای کل داده‌ها نمایش داده شده است. محور افقی اندیس داده‌ها و محور عمودی مقادیر دما هدف و پیش‌بینی شده نمایش داده شده است. شکل (۴) رگرسیون به ازای انواع داده‌ها در شبکه عصبی GMDH بهبود یافته با الگوریتم حشره آب‌سوار را نشان می دهد.

در شکل (۳-ب) خروجی های هدف با دایره‌های قرمز و خروجی‌های شبکه عصبی روش گروهی مدل‌سازی داده بهبود یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما با مربع‌های آبی رنگ به ازای داده‌های آزمایشی نمایش داده شده است. محور افقی اندیس داده‌ها و محور عمودی مقادیر دما هدف و پیش‌بینی شده نمایش داده شده است. در شکل (۳-ج) خروجی‌های هدف با دایره‌های قرمز و خروجی‌های



شکل ۳- خروجی‌های هدف و خروجی‌های GMDH بهبودیافته با الگوریتم حشره آب‌سوار به ازای داده‌های آموزشی (الف) و آزمایشی (ب) و کل داده‌ها (ج)



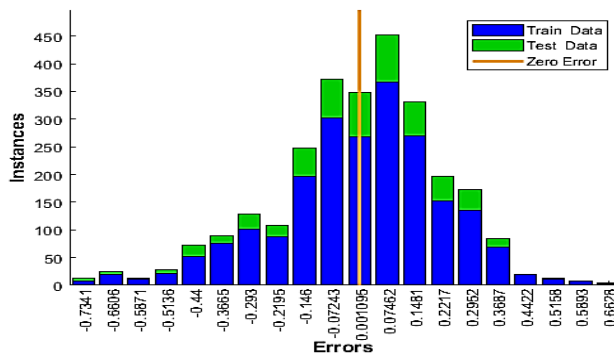
شکل ۴- نمودار رگرسیون به ازای انواع داده‌ها در شبکه عصبی GMDH بهبودیافته با الگوریتم حشره آب‌سوار

شده است محور افقی مقدار خطا و محور عمودی فراوانی خطا در شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب‌سوار را نشان می‌دهد. در شکل (۶) خروجی‌های هدف با دایره‌های قرمز و خروجی‌های شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب‌سوار در پیشبینی دما با مربع‌های آبی رنگ به ازای داده‌های آموزشی نمایش داده شده است. محور افقی اندیس داده‌ها و محور عمودی مقادیر دما هدف و پیشبینی شده نمایش داده شده است

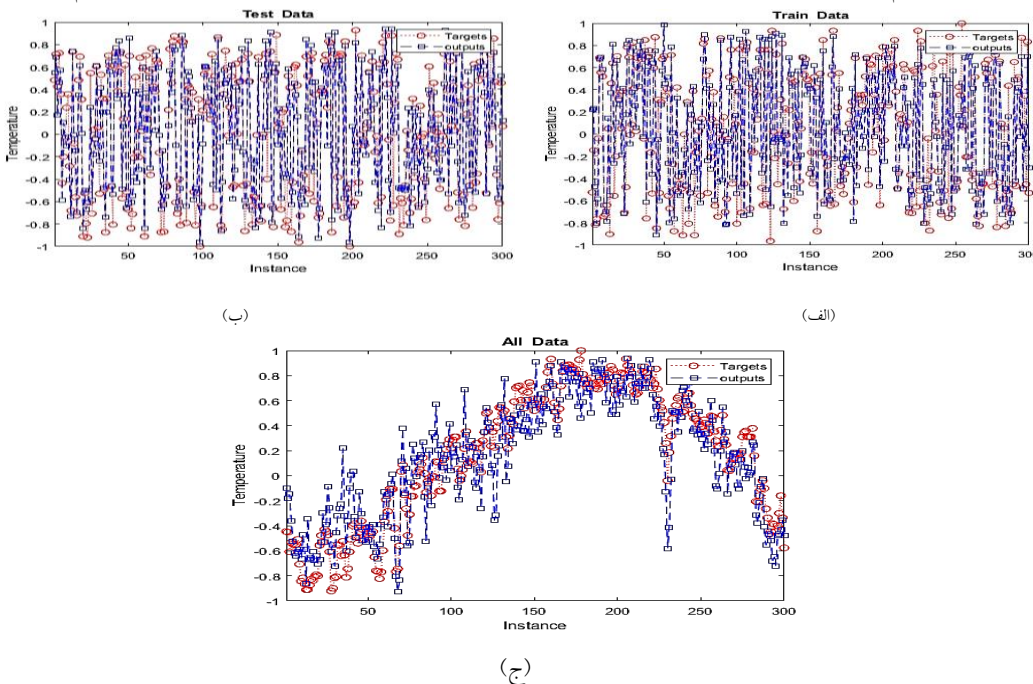
نتایج پیشبینی دما با شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب‌سوار در استفاده از شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب‌سوار در پیش‌بینی دما، تعداد نورون‌ها در لایه پنهان برابر ۵، نوع توابع فعال‌سازی در لایه‌های پنهان و خروجی تانزانت هاپربولیک در نظر گرفته شده است و نتایج بر اساس انواع خطا مطابق جدول (۵) است. در شکل (۵) هیستوگرام خطا به ازای داده‌های آموزشی با رنگ آبی و به ازای داده‌های آزمایشی با رنگ سبز نمایش داده

جدول ۵- انواع خطا در شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار در پیشبینی دما

	MSE	RMSE	MAE	SSE
Train Data	۰.۰۵۳۵	۰.۲۳۱۴	۰.۱۸۰۵	۱۱۶.۵۲
Test Data	۰.۰۵۰۲	۰.۲۲۴	۰.۱۷۲۹	۲۷.۳
All Data	۰.۰۵۲۹	۰.۲۲۹۹	۰.۱۷۹	۱۴۳.۸۲



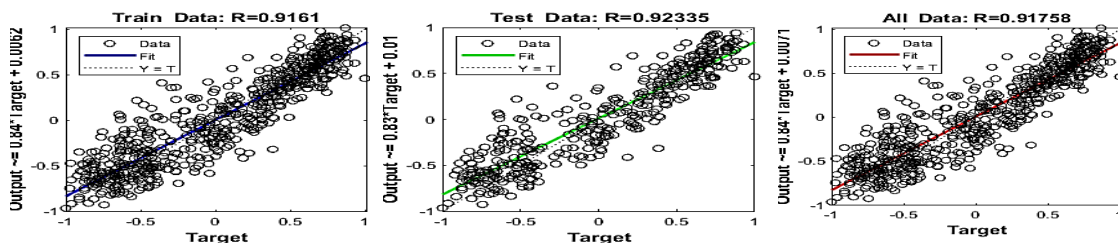
شکل ۵- هیستوگرام خطا به ازای داده‌های آموزشی و آزمایشی در شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار



شکل ۶- خروجی‌های هدف و خروجی‌های شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار به ازای داده‌های آموزشی (الف) و آزمایشی (ب) و کل داده‌ها (ج)

ازای داده‌های آزمایشی برابر ۰.۹۲۳ و به ازای کل داده‌ها برابر ۰.۹۱۸ است. هرچه مقدار ضریب رگرسیون به یک نزدیک‌تر باشد همبستگی بین خروجی‌های هدف و خروجی‌های مدل بیشتر و خطا در پیشبینی کم‌تر است.

شکل (۷) نمودار رگرسیون به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را در شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار را نشان می‌دهد. ضریب رگرسیون به ازای داده‌های آموزشی برابر ۰.۹۱۶، به



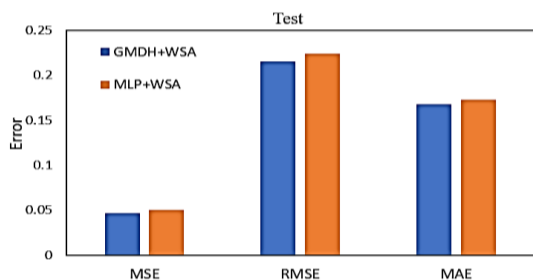
شکل ۷- نمودار رگرسیون به ازای انواع داده‌ها در شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار

در مقایسه با شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار دارد. با توجه به شکل (۸-ج) شبکه عصبی روش گروهی مدل‌سازی داده‌های بهبود یافته با الگوریتم حشره آب سوار از لحاظ انواع خطا به ازای کل داده‌ها عملکرد بهتری در پیش‌بینی دما در مقایسه با شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار دارد.

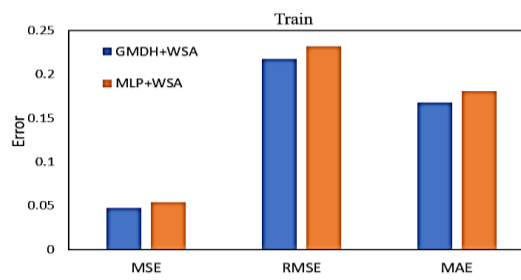
با توجه به جدول (۶) و شکل (۸-الف) شبکه عصبی روش گروهی مدل‌سازی داده‌های بهبود یافته با الگوریتم حشره آب سوار از لحاظ انواع خطا به ازای داده‌های آموزشی عملکرد بهتری در پیش‌بینی دما در مقایسه با شبکه عصبی پرسپترون چندلایه آموزش یافته با الگوریتم حشره آب سوار دارد. با توجه به شکل (۸-ب) شبکه عصبی روش گروهی مدل‌سازی داده‌های بهبود یافته با الگوریتم حشره آب سوار از لحاظ انواع خطا به ازای داده‌های آزمایشی عملکرد بهتری در پیش‌بینی دما

جدول ۶- انواع خطا در پیش‌بینی دما به ازای داده‌های آموزشی و آزمایشی و کل داده‌ها

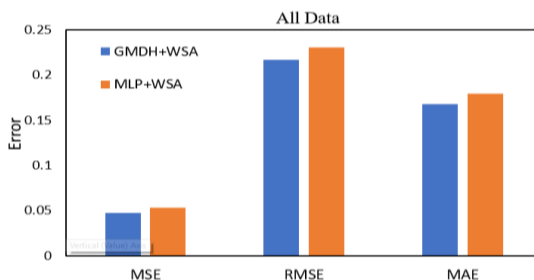
	داده‌های آموزشی		داده‌های آزمایشی		کل داده‌ها	
	MLP+WSA	GMDH+WSA	MLP+WSA	GMDH+WSA	MLP+WSA	GMDH+WSA
MSE	۰.۰۵۳۵	۰.۰۴۷۱	۰.۰۵۰۲	۰.۰۴۶۲	۰.۰۵۲۹	۰.۰۴۶۹
RMSE	۰.۲۳۱۴	۰.۲۱۷۱	۰.۲۲۴	۰.۲۱۴۹	۰.۲۲۹۹	۰.۲۱۶۶
MAE	۰.۱۸۰۵	۰.۱۶۷۳	۰.۱۷۲۹	۰.۱۶۷۴	۰.۱۷۹	۰.۱۶۷۳



(ب)



(الف)



(ج)

شکل ۸- انواع خطا در پیش‌بینی دما به ازای داده‌های آموزشی (الف) و آزمایشی (ب) و کل داده‌ها (ج)

نتیجه گیری

این مقاله مبتنی بر پیشبینی دما با استفاده از شبکه عصبی روش گروهی مدل‌سازی داده و الگوریتم فراابتکاری حشره آب‌سوار است. در بخش دوم پیشینه تحقیق مرور شد و در بخش سوم الگوریتم‌های حشره آب‌سوار و تحلیل مؤلفه‌های اصلی و شبکه‌های عصبی روش گروهی مدل‌سازی داده و پرسپترون چندلایه مرور شدند. در بخش چهارم به مراحل روش پیشنهادی در پیشبینی دما پرداخته شد و در بخش پنجم شبیه‌سازی مسأله انجام و نتایج کمی و کیفی استفاده از شبکه عصبی روش گروهی مدل‌سازی داده بهبودیافته با الگوریتم حشره آب‌سوار و شبکه عصبی پرسپترون چند لایه آموزش‌یافته با الگوریتم حشره آب‌سوار در پیشبینی دما نمایش داده شد. با توجه به جداول (۴) و (۵) شبکه عصبی روش گروهی مدل‌سازی داده بهبودیافته با الگوریتم حشره آب‌سوار و شبکه عصبی پرسپترون چند لایه آموزش‌یافته با الگوریتم حشره آب‌سوار به ترتیب با میانگین مربعات خطا 0.0529 و 0.0469 پیشبینی دما را انجام دادند. بنابراین بهترین عملکرد متعلق به شبکه عصبی روش گروهی مدل‌سازی داده بهبودیافته با الگوریتم حشره آب‌سوار و بدترین عملکرد متعلق به شبکه عصبی پرسپترون چندلایه آموزش‌یافته با الگوریتم حشره آب‌سوار است.

منابع

4. F. Tasnádi, F. Bock, J. Tidholm, A. V. Shapeev, and I. A. Abrikosov, "Efficient and accurate prediction of elastic properties of Ti0.5Al0.5N at elevated temperature using machine learning interatomic potential," *Thin Solid Films*, p. 138927, 2021.
5. I. Gupta, H. Mittal, D. Rikhari, and A. K. Singh, "MLRM: A Multiple Linear Regression based Model for Average Temperature Prediction of A Day," *arXiv preprint arXiv:2203.05835*, 2022.
6. J. Refonaa and M. Lakshmi, "Remote sensing based rain fall prediction using big data assisted integrated routing framework," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11, 2021.
7. Y. Cheng, Z. Zheng, J. Wang, L. Yang, and S. Wan, "Attribute reduction based on genetic algorithm for the coevolution of meteorological data in the industrial internet of things," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.
8. Z. Karevan and J. A. Suykens, "Transductive LSTM for time-series prediction: An application to weather forecasting," *Neural Networks*, vol. 125, pp. 1-9, 2020.
9. S. Dhamodharavadhani and R. Rathipriya, "Region-wise rainfall prediction using mapreduce-based exponential smoothing techniques," in *Advances in big data and cloud computing*: Springer, pp. 229-239, 2019.
10. E. Soares, P. Costa Jr, B. Costa, and D. Leite, "Ensemble of evolving data clouds and fuzzy models for weather time series prediction," *Applied Soft Computing*, vol. 64, pp. 445-453, 2018.
11. R. I. Rasel, N. Sultana, and P. Meesad, "An application of data mining and machine learning for weather forecasting," in *International conference on computing and information technology*, pp. 169-178: Springer, 2017.
12. S. Sahoo, "A Parallel Forecasting Approach Using Incremental K-means Clustering Technique," in *Computational Intelligence in Data Mining*: Springer, pp. 165-172, 2017.
13. A. C. Onal, O. B. Sezer, M. Ozbayoglu, and E. Dogdu, "Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 2037-2046: IEEE, 2017.
14. A. Kaveh and A. D. Eslamlou, "Water strider algorithm: A new metaheuristic and applications," in *Structures*, vol. 25: Elsevier, pp. 520-541, 2020.
1. B. Azari and M. Tabesh, "Urban storm water drainage system optimization using a sustainability index and LID/BMPs," *Sustainable Cities and Society*, vol. 76, p. 103500, 2022.
2. K. Azizi and C. I. Meier, "Urban Pluvial Flood Risk Assessment: Challenges and Opportunities for Improvement Using a Community-Based Approach," in *World Environmental and Water Resources Congress 2021*, pp. 350-361, 2021.
3. R. Palamuttam, R. Palamuttam, R. M. Mogrovejo, C. Mattmann, B. Wilson, K. Whitehall, R. Verma, L. McGibney and P. Ramirez, "SciSpark: Applying in-memory distributed computing to weather event detection and tracking," in *2015 IEEE international conference on Big data (Big data)*, pp. 2020-2026: IEEE, 2015.

18. B. Azari, K. Hassan, J. Pierce, and S. Ebrahimi, "Evaluation of machine learning methods application in temperature prediction," *Environ Eng*, vol. 8, pp. 1-12, 2022.
15. B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20-30, 2021.
16. J. Wu, Y. Wang, X. Zhang, and Z. Chen, "A novel state of health estimation method of Li-ion battery using group method of data handling," *Journal of Power Sources*, vol. 327, pp. 457-464, 2016 .
17. R. Vang-Mata, *Multilayer Perceptrons: Theory and Applications*. Nova Science Publishers, 2020.