

بهبود دقت در مدلسازی عمق برف مبتنی بر اثربخشی انتخاب ویژگی‌های شاخص در رگرسیون‌های پارامتریک و غیر پارامتریک

مصطفی آصفی^۱ و علی فتحزاده^{۲*}

چکیده

دستیابی به توزیع مکانی عمق برف می‌بایست از راه مشاهده‌ای و در مقیاسی فشرده صورت گیرد. لیکن به دلیل محدودیت‌های عملی، جمع‌آوری اطلاعات، به‌ویژه در مقیاس‌های مذکور، دشوار و گاهی غیرممکن است. با توجه به مشکلات موجود، استفاده از رویکرد یادگیری ماشین و انتخاب ویژگی می‌تواند سبب افزایش کاربردپذیری پهنه‌بندی عمق برف در مناطق مرتفع شود. در این پژوهش اثربخشی کاهش ویژگی‌های غیر مؤثر در یادگیری مبتنی بر مدل‌های پارامتریک و غیر پارامتریک بررسی شد. نمونه‌های استفاده‌شده برای بررسی فرضیه‌ها از منطقه چلگرد ایران برداشت شده است. به همین منظور ابتدا با استفاده از روش هایپرکیوب محل ۱۰۰ نقطه مشخص و طی یک عملیات صحرایی داده‌های عمق برف در نقاط مورد نظر و همچنین در ۱۹۵ نقطه دیگر به صورت تصادفی و با نمونه‌بردار مدل فدرال برداشت شد. سپس با استفاده از مدل رقومی ارتفاع، ۲۵ پارامتر ژئومورفومتری استخراج و همراه با ۶ باند تصاویر ماهواره لندست ۸ و شاخص NDSI به‌عنوان ورودی‌های مدل‌ها انتخاب شد. نتایج حاصل از این پژوهش نشان داد که روش‌های پارامتریک و ناپارامتریک دقت قابل‌قبولی در مدلسازی عمق برف نداشتند، اما مدل رگرسیون خطی با روش انتخاب ویژگی حریصانه پیش‌رو و بهینه‌ساز جمعیت ذره‌ها با میانگین مربعات خطا برابر با ۲۲/۱۷ و ۲۲/۱۹ توانستند با دقت بهتری تغییرات عمق برف را مدل کنند.

واژه‌های کلیدی: انتخاب ویژگی پیش‌رو، کاهش ویژگی، نمونه‌برداری، عمق برف، یادگیری ماشین.

ارجاع: آصفی م. و فتحزاده ع. ۱۴۰۱. بهبود دقت در مدلسازی عمق برف مبتنی بر اثربخشی انتخاب ویژگی‌های شاخص در رگرسیون‌های پارامتریک و غیر پارامتریک. مجله پژوهش آب ایران. ۴۷: ۵۷-۶۷. <https://dx.doi.org/10.22034/IWRJ.2022.14028.2445>

۱- دانش‌آموخته کارشناسی ارشد آبخیزداری، دانشگاه اردکان.

۲- دانشیار گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان.

* نویسنده مسئول: fat@ardakan.ac.ir

تاریخ پذیرش: ۱۴۰۱/۰۹/۰۵

تاریخ دریافت: ۱۴۰۱/۰۶/۲۶

مقدمه

یکی از آشکارترین ویژگی‌های پوشش برفی کوهستانی، ناهمگنی مکانی آن است (سلیگمان، ۱۹۳۶؛ مک کی و گری، ۱۹۸۱؛ پومری و گری، ۱۹۹۵). این ناپایداری مکانی، در سطح وسیعی از مقیاس‌ها وجود دارد که از مقیاس کمتر از متر تا صدها کیلومتر رده‌بندی می‌شود. ناهمگنی پوشش برفی کوهستان، تأثیر شگرفی بر شکل‌گیری برف (اسچوایزر و همکاران، ۲۰۰۳) و آب‌شناسی دارد (بالک و الدر، ۲۰۰۰؛ لاندکوئیست و دینگر، ۲۰۰۵؛ باوای و همکاران، ۲۰۱۳). با توجه به پیچیدگی فرایند برآورد پارامترهای عمق برف و مشکلات موجود در اندازه‌گیری پارامترهای برف در مناطق کوهستانی و همچنین در مناطق مرتفع استفاده از روش‌های مختلف مبتنی بر هوش مصنوعی و یادگیری ماشین توصیه می‌شود (طبری و همکاران، ۱۳۸۸).

اساس روش‌های یادگیری ماشین، در واقع یادگیری الگوهای ذاتی بین مقادیر ویژگی‌ها و ارتباطات بین آن‌ها و عمق برف است که می‌تواند برای پیش‌بینی عمق برف در مکانی جدید، استفاده شود (پاندى، ۲۰۱۳). تاکنون متخصصان و کارشناسان زیادی، از روش‌های یادگیری ماشین مختلفی در این زمینه استفاده کرده و به دقت‌های قابل توجهی دست یافته‌اند (اسوار و گوپتا، ۲۰۱۳؛ پاندى، ۲۰۱۳؛ الیش و همکاران، ۲۰۱۳؛ کایرساپ و همکاران، ۲۰۰۲؛ پاپازچاروس و همکاران، ۲۰۱۲). قرایی منش و همکاران (۲۰۱۶)، مدلسازی توزیع مکانی عمق برف در یک منطقه نیمه‌خشک در ایران را با استفاده از شبکه عصبی مصنوعی و درخت تصمیم انجام دادند و نتایج حاکی از عملکرد بهتر مدل درخت تصمیم بود.

با توجه به عدم امکان نمونه‌برداری عمق برف در ارتفاعات و محل‌های صعب‌العبور، استفاده از روش‌های یادگیری ماشین، می‌تواند نقش بسزایی در کاربردپذیری پیش‌بینی عمق برف داشته باشد. مزیت‌های این نوع از روش‌های تخمین، شامل توانایی مدل‌های هوش محاسباتی برای مدل‌کردن مجموعه‌های پیچیده از ارتباطات بین عمق برف و ویژگی‌های محل برف با یادگیری از داده‌های مشابه است (الیش و همکاران، ۲۰۱۳).

مدل‌های تخمین عمق برف از مجموعه بزرگی از ویژگی‌ها که تعیین‌کنندگان عمق نامیده می‌شوند، برای تخمین استفاده می‌کنند. همه این ویژگی‌ها برای برآوردی دقیق

تأثیرگذار نیستند؛ از این رو در حوزه تخمین عمق برف، الگوریتم‌های انتخاب زیرمجموعه‌ای از پارامترها، برای بهینه‌کردن دقت مدل‌ها برای نمونه‌های دیده‌نشده، با هدف انتخاب زیر مجموعه‌ای از مؤثرترین پارامترها در برآورد عمق برف، مورد استفاده قرار گرفته‌اند (پاپازچاروس و همکاران، ۲۰۱۲). با توجه به اهمیت فراوان تغییرات عمق برف و برآورد دقیق آن تاکنون از روش‌های زیادی چون زمین آمار، هوش مصنوعی و تصاویر راداری (حق‌زاده و همکاران، ۲۰۱۷؛ تدسکو و همکاران، ۲۰۰۴ و تسی و همکاران، ۲۰۱۹) استفاده شده است. ضرورت محاسبه دقیق‌تر عمق برف می‌طلبد که از روش‌ها، پارامترها و همچنین یک نمونه‌برداری مناسب که گنج خانلو و همکاران (۲۰۲۰)، تکنیک هایپرکیوب را دقیق‌تر دانستند استفاده شود؛ بنابراین در این تحقیق از روش‌های پارامتریک، ناپارامتریک و کاهش ویژگی و همچنین از پارامترهای ژئومورفومتری و تصاویر ماهواره‌ای به‌عنوان ورودی‌ها در برآورد عمق برف بهره گرفته شد و نتایج مقایسه شد.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

منطقه چلگرد واقع در شهرستان کوهرنگ است که در منطقه غربی استان چهارمحال و بختیاری و کشور ایران است. این منطقه در حد فاصل طول جغرافیایی "۲۹°۵۶' ۴۹° تا "۳۷°۲۶' ۵۰° شرقی و عرض جغرافیایی "۵۴°۵۹' ۳۱° تا "۵°۴۹' ۳۲° قرار دارد؛ بنابراین مشخصات اقلیمی این ناحیه به این صورت است که در متوسط دمای سالانه ۹ درجه سانتی‌گراد، بارش ۱۳۰۰ میلی‌متر، طول دوره یخبندان ۱۳۰ روزه و رژیم بارش زمستانه قرار گرفته است (حیدری و همکاران، ۱۳۹۷).

نمونه‌برداری صحرائی

در این پژوهش ابتدا بررسی‌های اولیه نظیر راه‌های دسترسی، ترتیب نمونه‌برداری و محدودیت‌های نمونه‌برداری انجام و سپس نسبت به تهیه اطلاعات مورد نیاز و محدوده مطالعاتی اقدام شد. در ادامه با به‌کارگیری تکنیک ابرمکعب لاتین یا هایپرکیوب^۱ به‌عنوان یک روش

1- Hypercube

روش‌های یادگیری ماشین^۴

روش‌های یادگیری ماشین، الگوهایی را از داده‌های گذشته یاد می‌گیرند و از این الگوها برای پیش‌بینی بخشی از ویژگی‌های نمونه‌های جدید استفاده می‌کنند. ایده‌آسی استفاده از روش‌های یادگیری ماشین برای تخمین و پیش‌بینی عمق برف، این است که دادگان قدیمی، شامل نمونه‌های مشاهده‌شده زیادی هستند که با ویژگی‌های با ارزششان برای توصیف هر محل نمونه‌برداری، تشریح شده‌اند. محل‌های پوشش برف با ویژگی‌های مشابه، ممکن است شامل عمق‌های مشابهی نیز باشند. کار روش‌های یادگیری ماشین، یادگیری الگوهای ذاتی بین مقادیر ویژگی‌ها و ارتباطات بین آن‌ها و عمق برف است که می‌تواند برای پیش‌بینی عمق برف در مکانی جدید، استفاده شود (پاندی، ۲۰۱۳).

از این‌رو در این پژوهش سعی شده است با استفاده از روش‌های یادگیری ماشین برای انجام تخمینی هرچه دقیق‌تر، روشی نو با استفاده از پیاده‌سازی روش انتخاب ویژگی بسته‌بندی براساس رویکرد حرصانه پیش‌رو و آنالیز مؤلفه اصلی مبتنی بر مدل‌های پارامتریک و غیرپارامتریک صورت پذیرد.

مدل‌سازی به روش پارامتریک

تنها روشی که در این مدل‌سازی استفاده شد، روش رگرسیون خطی چندمتغیره است. در روش رگرسیون خطی چندمتغیره، فرض می‌شود که پارامتر حقیقی r به صورت یک تابع خطی (یعنی یک مجموع وزن‌دار) از چندین متغیر ورودی x_1, \dots, x_d به همراه «یک پارامتر نویز» نوشته می‌شود. منظور از پارامتر نویز، متغیرهای اضافی است که ما قادر به مشاهده آن‌ها نیستیم. معادله (۳) مدل رگرسیون خطی چندمتغیره را نشان می‌دهد.

$$r^t = g(x^t | w_0, w_1, \dots, w_d) + \epsilon w_0 + w_1 x_1^t + \dots + w_d x_d^t \epsilon, \quad (3)$$

که در آن r^t عبارت است از پارامتر خروجی مورد نظر با شاخص t پارامتر بهینه‌شده در بهترین برازش در داده‌های آموزش و ϵ پارامتر نویز با توزیع نرمال با میانگین صفر و واریانس ثابت است. در معادله فوق، می‌توان نشان داد که بیشینه‌کردن درست‌نمایی^۵ معادل با کمینه‌کردن

نمونه‌برداری تصادفی و با استفاده از نقشه‌های ارتفاع، شیب و جهت دامنه محل ۱۰۰ نقطه در جهت نمونه‌برداری در منطقه جلگرد انتخاب شد و مختصات جغرافیایی هریک از نقاط پس از تبدیل فرمت به وسیله نرم‌افزار Global Mapper و به کارگیری نرم‌افزار Map source به GPS منتقل شد. تکنیک هایپرکیوب یک روش مدرن است که به‌ندرت در نمونه‌برداری از عمق برف استفاده شده است که در آن نمونه مورد نظر به تعداد n از متغیرهای چندگانه به‌شکلی تعیین می‌شود که نمونه مرتبط با هریک از متغیرها از حداکثر طبقه‌بندی برخوردار باشد (مک‌کی و همکاران، ۱۹۷۹). در ادامه طی یک عملیات میدانی و پیمایش برف سه روزه ۲۷، ۲۸ و ۲۹ اسفند ۱۳۹۲ و به‌وسیله نمونه‌برداری مدل مونت رز عمق برف در هریک از نقاط برداشت شد. علاوه بر این نقاط، ۱۹۵ نقطه دیگر نیز به‌طور تصادفی از منطقه مورد مطالعه جمع‌آوری شد. بعد از برداشت نمونه‌های برف، سایر نقشه‌ها و اطلاعات مورد نیاز برای پهنه‌بندی که شامل داده‌های مرتبط با تصاویر ماهواره‌ای و پارامترهای مشتق‌شده از مدل رقومی ارتفاع است، در محیط نرم‌افزار ساگا (SAGA) استخراج و در جدول ۱ آورده شده است. مدل رقومی استفاده‌شده با اندازه پیکسل ۱۰ متر است. همچنین تصاویر مورد استفاده، تصاویر ماهواره‌ای لندست ۸ است که قبل از استفاده از این تصاویر و همچنین مدل رقومی ارتفاع تصحیحات لازم اعمال شد. در جهت نرمال کردن داده‌ها نیز از روش بزرگ‌ترین-کوچک‌ترین^۲ استفاده شد و مقادیر ویژگی‌ها بین صفر و یک نرمال شدند که در معادله (۱) قابل مشاهده است.

$$\text{MinMax-normalization} = \frac{x_i - \text{Min}(Att_i)}{\text{Max}(Att_i) - \text{Min}(Att_i)} \quad (1)$$

در جهت محاسبه شاخص NDSI^۳ توسط باندهای لندست ۸ از معادله (۲) استفاده شد:

$$\text{Landsat 8 NDSI} = \frac{\text{Band3} - \text{Band6}}{\text{Band3} + \text{Band6}} \quad (2)$$

1- System for Automated Geoscientific Analyses

2- Max-Min Normalization

3- Normalized Difference Snow Index

4- Machine Learning

5- Maximization of likelihood

مجموع مربعات خطای معادله (۴) است.

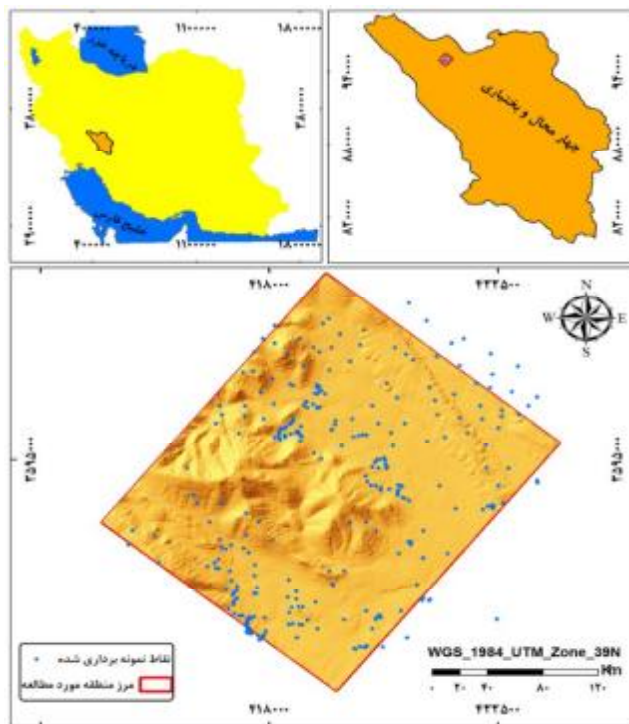
$$E(w_0, w_1, \dots, w_d | X) = \frac{1}{2} \sum_t (r^t - w_0 - w_1 x_1^t - \dots - w_d x_d^t)^2, \quad (4)$$

که با مشتق گرفتن از این رابطه و همچنین بازنویسی عبارات منتج شده از مشتق گیری به فرم برداری و ماتریسی

در نهایت به معادله (۵) برای مدل رگرسیون خطی

چندمتغیره می‌رسیم.

$$X^t X w = X^t r \Rightarrow w = (X^T X)^{-1} X^T r. \quad (5)$$



شکل ۱- موقعیت مکانی منطقه مورد مطالعه و نقاط نمونه برداری شده در استان چهارمحال و بختیاری

جدول ۱- پارامترهای کمکی به کاررفته در مدل

پارامترهای مورد استفاده		
ارتفاع نرمال شده	شیب حوضه آبخیز	باندهای (۲.۳.۴.۵.۶.۷)
ارتفاع استاندارد شده	مساحت اصلاح شده حوضه آبخیز	شیب
موقعیت میانی شیب	طول شیب	جهت جغرافیایی
شاخص همواری قله	شاخص قدرت آبراهه	شاخص همگرایی
شاخص همواری دره	نیمرخ عرضی انحنای	سطح ویژه حوضه
شاخص تفاوت نرمال شده برف	نیمرخ طولی انحنای	شاخص رطوبتی توپوگرافی
عمق چاله‌ها	انحنای عمومی	فاکتور طول شیب
ارتفاع حوضه	اثر باد	سطح اساس شبکه زهکشی
مدل رقومی ارتفاع	ارتفاع شیب	تراکم جریان

مدل سازی به روش غیر پارامتریک

در روش‌های پارامتریک فرض بر این است که یک مدل مشخص روی همه نمونه‌ها قابل پیاده‌سازی است، از مزایای این فرضیه این است که مسئله تخمین را به تعدادی پارامترهای کمی تبدیل می‌کنیم، اما این فرضیه همیشه نمی‌تواند درست باشد و همین امر باعث می‌شود که در

صورت درست نبودن آن، با یک خطای بالا مواجه شویم؛ بنابراین باید توابعی داشته باشیم که با خطای کمتری، تخمین را برای ما انجام دهد (آلپادین، ۲۰۱۴). روش‌های غیر پارامتریک به جای برازش دادن یک سری توابع مشخص بر داده‌ها، بر اساس تشخیص الگو و اصل تشابهات عمل می‌کند و در مواردی که ارتباط بین ورودی و خروجی از

رابطه مورد استفاده برای این دسته از روش‌ها به شکل معادله (۸) است. در این معادل، $k(u)$ بیانگر هسته مورد استفاده است. مقدار h در این معادله به صورت ثابت در نظر گرفته می‌شود (آلپادین، ۲۰۱۴).

$$\hat{g}(x) = \frac{\sum_{t=1}^N k\left(\frac{x-x^t}{h}\right)r^t}{\sum_{t=1}^N k\left(\frac{x-x^t}{h}\right)} \quad (۸)$$

رگرسیون k-همسایه نزدیک‌تر^۷

مدل نزدیک‌ترین همسایگی، یکی از روش‌های داده‌کاوی است که هدف کلی آن طبقه‌بندی و تخمین ویژگی‌های یک سری داده‌های مجهول با توجه به بیش‌ترین شباهت این داده‌ها با داده‌های معلومی است که در همسایگی (نزدیکی) آن‌ها قرار دارند (فدایی کرمانی و همکاران، ۱۳۹۳). روال کار این تخمین‌گر به این صورت است که به ازای هر نمونه x از داده‌های تست، فاصله آن نمونه با همه نمونه‌های داده‌های آموزش محاسبه می‌شود. k نمونه‌ای که فاصله کمتری با نمونه x داشته باشد، انتخاب شده و کلاس این k نمونه استخراج می‌شود. هر کلاسی که بیشترین تکرار را میان این k نمونه داشته باشد، به عنوان کلاس پیش‌بینی شده برای نمونه x در نظر گرفته می‌شود. برای بالابردن کارایی این تخمین‌گر باید بهترین k انتخاب شود. روال انتخاب بهترین k همانند دو تخمین‌گر قبل است، در این تخمین‌گر $P(c_i|x)$ به صورت معادله (۹) محاسبه می‌شود:

$$P(c_i|x) = \frac{k_i}{k} \quad (۹)$$

k_i نمایانگر تعداد نمونه‌هایی از k نمونه انتخابی که عضو کلاس i هستند و k نمایانگر تعداد نمونه‌هایی که کمترین فاصله با نمونه x دارند، است.

مدل‌سازی به روش کاهش ویژگی

چالشی که روش‌های یادگیری ماشین با آن مواجه هستند، این است که بسیار به ویژگی‌های تشریح‌کننده نمونه‌ها حساسند. دادگانی که برای آموزش این روش‌ها مورد استفاده قرار می‌گیرند، معمولاً دربرگیرنده ویژگی‌هایی هستند که نه تنها تأثیری بر افزایش دقت این روش‌ها ندارند، بلکه باعث کاهش دقت آن‌ها نیز می‌شوند. به این

قبل به‌طور کامل مشخص نباشد، مؤثر خواهد بود (شارما و اونیل، ۲۰۰۲ و یاکوویتز، ۱۹۹۳). برای ارزیابی روش‌های غیرپارامتریک در حل مسئله رگرسیون چندمتغیره، سه روش رگرسیون مبتنی بر میانگین^۱، رگرسیون مبتنی بر هسته^۲ و رگرسیون K-همسایه نزدیک‌تر^۳ مورد استفاده قرار گرفته‌اند که در ادامه به معرفی آن‌ها می‌پردازیم.

رگرسیون مبتنی بر میانگین^۴

در روش‌های غیرپارامتریک، نمونه‌های آموزشی براساس شباهتی که بین‌شان برقرار است، در دسته‌هایی^۵ تقسیم می‌شوند. به هنگام ساختن مدل برای تخمین مقدار هدف نمونه مورد آزمایش، شبیه‌ترین دسته به نمونه به میزان فاصله h معینی مشخص می‌شود و نمونه مورد نظر به آن دسته اختصاص می‌یابد. در ادامه کار اگر از رگرسیون مبتنی بر میانگین برای عمل تخمین استفاده شود، میانگین مقدار هدف نمونه‌های آموزشی دسته‌ای که نمونه به آن اختصاص یافته، به عنوان مقدار هدف آن نمونه در نظر گرفته می‌شود. این روش به صورت فرمولی بیان شده است که در این فرمول $\hat{g}(x)$ میانگین دسته‌ای که براساس $w(u)$ نمونه مورد آزمایش به آن اختصاص یافته را محاسبه می‌کند و t بیانگر نمونه‌های آموزشی مشابه نمونه مورد آزمایش است (آلپادین، ۲۰۱۴).

$$\hat{g}(x) = \frac{\sum_{t=1}^N w\left(\frac{x-x^t}{h}\right)r^t}{\sum_{t=1}^N w\left(\frac{x-x^t}{h}\right)} \quad (۶)$$

$$w(u) = \begin{cases} 1 & \text{if } |u| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (۷)$$

رگرسیون مبتنی بر هسته^۶

این نوع از روش‌های یادگیری غیرپارامتریک هم مانند روش رگرسیون مبتنی بر میانگین، از شباهت بین نمونه‌های آموزشی و آزمون برای تخمین استفاده می‌کنند. با این تفاوت که به جای تابع فاصله از یک هسته برای بررسی شباهت استفاده می‌کند. هسته گاوسی یکی از نمونه هسته‌های پرکاربرد در این زمینه محسوب می‌شود.

1- Mean Smoother
2- Kernel Smoother
3- k-Nearest Neighbor
4- Mean Smoother
5- Bin
6- Kernel Smoother

7- K-Nearest Neighbor

همسایگی جی‌بست^۵ استفاده می‌شود. یعنی هر ذره بهترین موقعیت قبلی خود و بهترین موقعیت قبلی هریک از ذرات موجود در جمعیت را به خاطر می‌آورد و به عبارت دیگر هر ذره درجهت بهترین موقعیت قبلی خود و به سمت بهترین ذره در حرکت است. هر ذره نماینده یک جواب ممکن است که به صورت تصادفی در فضای مسئله حرکت می‌کند (پولی و همکاران، ۲۰۰۷ و وانگ و همکاران، ۲۰۱۵).

سنجش و ارزیابی

بدین منظور با کمک روش‌های پارامتریک، ناپارامتریک و کاهش ویژگی مدل‌شده عمق برف در نقاط مورد نظر بررسی شد و سپس مجذور میانگین مربعات خطا (RMSE)، مقدار میانگین خطای نسبی (MMRE^۶)، و مقدار میانه خطای نسبی (MdMRE) درجهت ارزیابی دقت مدل‌ها محاسبه شدند.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Actual_i - Estimated_i)^2} \quad (10)$$

که در آن N به معنای تعداد نمونه‌های تست‌دادگان، $Actual_i$ به معنای مقدار واقعی نمونه i -ام، و $Estimated_i$ به معنای مقدار تخمین‌زده شده برای نمونه i -ام است.

$$MMRE = \frac{100}{N} \sum_{i=1}^N MRE_i \quad (11)$$

$$MRE_i = \frac{|Actual_i - Estimated_i|}{Actual_i} \quad (12)$$

$$MdMRE = 100 \times Median(MRE) \quad (13)$$

$$MRE = \{MRE_1, MRE_2, \dots, MRE_N\} \quad (14)$$

در این معادلات MRE با استفاده از معادله (۱۴) و MRE_i با استفاده از معادله (۱۲) محاسبه می‌شود.

نتایج و بحث

در این پژوهش، از روش‌های مختلف پارامتریک و غیرپارامتریک یادگیری ماشین برای تخمین عمق برف منطقه چلگرد شهرستان کوه‌رنگ استفاده شد. براساس اینکه دادگانی که برای آموزش این روش‌ها استفاده

دسته از ویژگی‌ها، ویژگی‌های زائد گفته می‌شود؛ از این رو روش‌های مختلف کاهش بُعد برای کاهش تأثیر این ویژگی‌ها ارائه شده‌اند (آلپادین، ۲۰۱۴). در این پژوهش درجهت کاهش بُعد از سه روش استخراج ویژگی با روش PCA و انتخاب ویژگی با روش پیش‌رو^۱ و الگوریتم ازدحام ذرات استفاده شده است که در ادامه تشریح شده‌اند.

کاهش ویژگی با روش PCA^۲

در الگوریتم PCA هدف این است که واریانس نمونه‌ها را ماکزیم کنیم. اگر نمونه‌ای در فضای دو بُعدی باشد و ابعاد به گونه‌ای تغییر کند که این نمونه در فضای یک بُعدی پخش شود و واریانس آن‌ها بیشتر شود، ممکن است الگوریتم‌های یادگیری دیگر بهتر پاسخ دهند. به منظور حل این مسئله از قضیه لاگرانژ استفاده شد. برای این منظور هدف به دست آوردن بردار ویژه و مقدار ویژه ماتریس کواریانس است؛ بنابراین بردار ویژه‌ای که متناظر با بزرگ‌ترین مقدار ویژه باشد، انتخاب شده و در صورتی که آن را در تک تک نمونه‌ها ضرب کنیم، همه آن‌ها به فضای یک بُعدی منتقل شده و کواریانس بیشتر می‌شود.

کاهش ویژگی با روش پیش‌رو^۳

در این روش درجهت کاهش ویژگی ابتدا از یک ویژگی شروع و هر بار ویژگی‌ها اضافه می‌شود تا زمانی که یا تأثیری در دقت نداشته باشند، یا اینکه دقت را کاهش دهند (پوگینی و مک‌لون، ۲۰۱۷).

الگوریتم بهینه‌سازی ازدحام ذرات^۴

الگوریتم بهینه‌سازی ازدحام ذرات یک تکنیک بهینه‌سازی تصادفی بر مبنای جمعیت است که با الهام از پرواز اجتماعی پرندگان، شنای دسته‌جمعی ماهی‌ها و زندگی اجتماعی آن‌ها مدل شده است (کندی و ابرهات، ۱۹۹۵) و با استفاده از یک سری روابط ساده ترکیب‌بندی شده است. جمعیت (حرکت تجمعی) در PSO، شامل مجموعه‌ای از اعضاست که به هر عضو داخل جمعیت، یک ذره گفته می‌شود. در این تکنیک از مفهوم جانمایی

1- Forward-Selection

2- Principal Component Analysis

3- Forward-Selection

4- Particle Swarm Optimization (PSO)

5- Gbest Neighborhood Topology

6- Root Mean Square Error (RMSE)

7- Mean Magnitude of Relative Error (MMRE)

نتایج حاصل از پیاده‌سازی روش‌های رگرسیون مبتنی بر میانگین، رگرسیون مبتنی بر هسته و رگرسیون K- همسایه نزدیک‌تر روی دادگان مورد نظر با تقسیم‌بندی داده‌ها به‌صورت: ۸۰ درصد داده آموزشی و ۲۰ درصد داده تست، برای $h = 1/55$ در روش رگرسیون مبتنی بر میانگین، $h = 0/04$ در روش رگرسیون مبتنی بر هسته و $k =$ در روش K- همسایه نزدیک‌تر در جدول ۳ آمده است. در این جدول، نتایج این روش‌ها براساس سه معیار RMSE، MMRE و MdMRE مشاهده می‌شود. انتخاب این مقادیر برای پارامترهای h و k با استفاده از روش اعتبارسنجی متقابل صورت گرفته است. اعتبار متقابل روشی است برای محاسبه ناریب از خطای پیش‌بینی و بیشتر برای برآوردگرهایی استفاده می‌شود که دارای توزیع پیچیده هستند. در این روش ابتدا یک مشاهده را حذف کرده و مقدار آماده را برای $n-1$ مشاهده باقیمانده محاسبه می‌کنند و به کمک آن مقدار خطای پیش‌بینی حساب می‌شود. با ادامه‌دادن به این روند، برای سایر مشاهدات و میانگین گرفتن از آن‌ها برآورد خطای متقابل به‌دست می‌آید.

شده‌اند، دارای ویژگی‌هایی با محدوده‌های مختلف هستند، در ابتدای امر این دادگان را با استفاده از روش نرمال‌سازی بزرگ‌ترین-کوچک‌ترین، نرمال کرده و سپس روش‌های معرفی شده را مورد استفاده قرار دادیم. نتایج حاصل از پیاده‌سازی هریک از روش‌ها در ادامه آمده است.

نتایج حاصل از روش پارامتریک

نتایج پیاده‌سازی روش رگرسیون خطی روی دادگان مورد نظر با تقسیم‌بندی داده‌ها به‌صورت: ۸۰ درصد داده آموزشی و ۲۰ درصد داده تست در جدول ۲ آمده است. در این جدول، نتایج روش رگرسیون خطی براساس سه معیار RMSE، MMRE، MdMRE مشاهده می‌شود.

جدول ۲- نتایج به‌دست آمده از روش رگرسیون خطی روی

دادگان			
MdMRE	MMRE	RMSE	روش نرمال‌سازی
۳۸۵۷۲۰	۴۹۰۰۵۰	۲۷۱۳۵۰	نرمال‌سازی بزرگ‌ترین-کوچک‌ترین
			کوچک‌ترین

نتایج روش‌های ناپارامتریک

جدول ۳- نتایج به‌دست آمده از روش‌های ناپارامتریک

MdMRE	MMRE	RMSE	روش نرمال‌سازی	روش
۵۳/۶۴	۵۳/۴	۴۶/۶۷	نرمال‌سازی بزرگ‌ترین-کوچک‌ترین	رگرسیون مبتنی بر میانگین
۳۹/۷۱	۴۱/۵۹	۳۹/۲۹	نرمال‌سازی بزرگ‌ترین-کوچک‌ترین	رگرسیون مبتنی بر هسته
۴۳/۳۴	۴۱/۰۲	۳۷/۵۷	نرمال‌سازی بزرگ‌ترین-کوچک‌ترین	رگرسیون k- همسایه نزدیک‌تر

نتایج کاهش ویژگی‌ها

در این پژوهش درجهت استخراج و کاهش ویژگی‌ها از سه روش تجزیه مؤلفه اصلی، حریصانه پیشرو و الگوریتم ازدحام ذرات استفاده شد که نتایج حاصل از پیاده‌سازی این روش‌ها روی روش‌های پارامتریک و ناپارامتریک در جدول ۴، آورده شده است.

با توجه به روش‌های مورد استفاده در این پژوهش، مشخص شد روش پارامتریک که در آن رگرسیون خطی پیاده‌سازی شده بود، نتیجه مناسبی نداشت و میزان جذر میانگین مربعات خطای ۲۷۱۳۵۰ را ارائه کرد که نسبت به پژوهش‌های انجام شده در زمینه برف قابل قبول نیست؛ زیرا قرایی منش و همکاران (۱۳۹۲) در مطالعه‌ای به مقایسه روش شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره در مدل‌سازی عمق برف پرداختند و نشان دادند که مدل

رگرسیونی ۷۴ درصد از تغییرات عمق برف را با میزان جذر میانگین مربعات خطای ۱۷ محاسبه کرده و شبکه عصبی با آرایش ۱-۱۱-۳۰، ۸۵ درصد از تغییرات عمق برف را با میانگین مربعات خطای ۱۲/۱ حساب کرده است و همچنین طبری و همکاران (۱۳۸۸) به مقایسه سه روش رگرسیون غیرخطی، شبکه عصبی مصنوعی و روش تلفیق شبکه عصبی-الگوریتم ژنتیک درجهت برآورد ضخامت برف و آب معادل برف پرداختند. از میان روش‌های به‌کاررفته، روش شبکه عصبی تلفیق‌یافته با الگوریتم ژنتیک بهترین نتیجه را دربرداشته است؛ اما روش‌های ناپارامتریک نسبت به روش‌های پارامتریک نتایج قابل قبول‌تری را ارائه کردند؛ به‌طوری‌که روش K- همسایه نزدیک‌تر کمترین میزان میانگین مربعات خطا را دارد. دادگانی که برای آموزش این روش‌ها مورد استفاده قرار

می‌گیرند، معمولاً دربرگیرنده ویژگی‌هایی هستند که نه تنها تأثیری بر افزایش دقت این روش‌ها ندارند، بلکه باعث کاهش دقت آن‌ها نیز می‌شوند؛ بنابراین درجهت کاهش و استخراج ویژگی‌ها از روش‌های تجزیه مؤلفه اصلی، حریصانه پیشرو و الگوریتم ازدحام ذرات استفاده شد و کمترین میزان جذر میانگین مربعات خطا را روش‌های حریصانه پیشرو و ازدحام ذرات به دست آوردند. شکل ۲ رابطه همبستگی مقادیر عمق برف برآوردشده با استفاده از روش‌های مذکور را نسبت به عمق برف اندازه‌گیری شده نشان می‌دهد. با توجه به اینکه دو روش حریصانه پیشرو و ازدحام ذرات کمترین میزان خطا را داشتند، اقدام به تهیه نقشه پهنه‌بندی عمق برف با استفاده از این دو روش شد، (شکل ۳). همان‌گونه که مشاهده می‌شود، در هر دو روش بیشترین میزان عمق برف در مناطق غربی منطقه مورد مطالعه است که علت آن را می‌توان وجود ارتفاعات و بادهای غالب با جهت شرقی-غربی دانست. تأثیر جهت روی عمق برف به‌ازای جهت‌های مختلف، متفاوت است. با توجه به نتایج به‌دست‌آمده می‌توان به این موضوع پی برد که جهت‌های غربی به‌دلیل بادپناهی عمق برف بیشتر و جهت‌های شرقی به‌دلیل بادرویی عمق برف کمتری دارند که این نتایج همسو با نتایج گنج خانلو و همکاران (۱۳۹۶) و مارچاند و کلینگویت (۲۰۰۱) است. به‌گونه‌ای که آن‌ها ذکر کرده‌اند که در شیب‌های رو به باد، همبستگی بیشتری بین عمق برف و جهت شیب برقرار است. قرایی منش و همکاران (۲۰۱۶)، معتقدند که پارامتر انحنای ارتفاعات و دامنه‌ها در تله‌اندازی و تجمع برف تأثیرگذار است که شامل انحنای مسطح و قائم است. خط‌الفرهای موجود در ارتفاعات که قاعدتاً محل مناسبی برای انباشت برف است، اما رابطه معکوس عمق برف و عمق دره در حوضه مورد مطالعه مشهود است که نتایج آن حاکی از تأثیر تابش خورشید و شیب بر خط‌الرأس ارتفاعات است که باعث ذوب، ایجاد رواناب و کاهش عمق برف در این مناطق شده است. کم‌بودن عمق برف در مناطق شرقی و تا حدودی جنوبی و شمالی، علاوه بر وجود بادهای غالب منطقه می‌تواند تحت‌تأثیر کاربری کشاورزی و مسکونی مناطق پست و هموار، رابطه معکوس ارتفاع و دمای سطح و همچنین رابطه مستقیم سطح پوشش برف و ارتفاع (حلبیان و صلحی، ۱۳۹۹؛ اصغری سراسکانرود و همکاران،

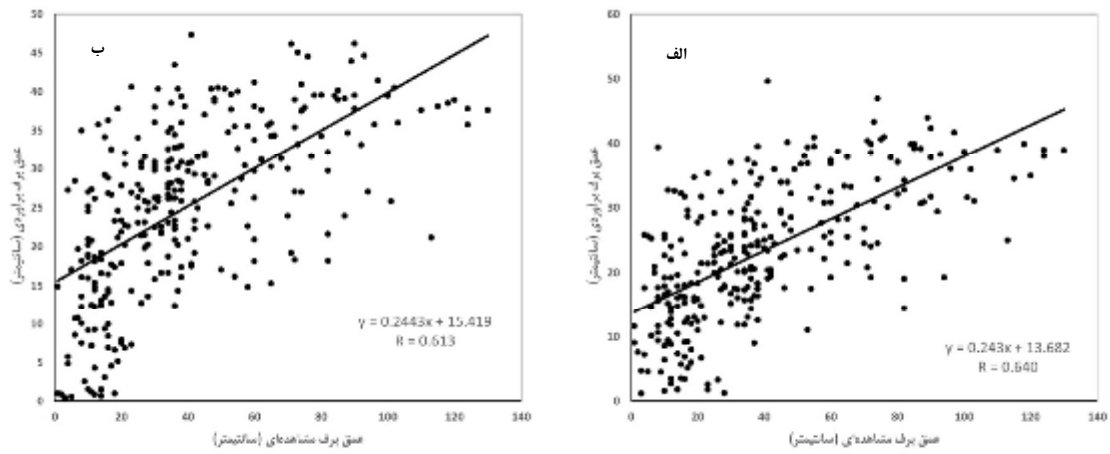
۱۴۰۰) باشد. رابطه عمق برف با ارتفاع در حوضه مورد مطالعه، متأثر از ترازهای ارتفاعی است. عمق برف این حوضه در ارتفاعات پایین کمتر و در ارتفاعات بالاتر بیشتر است. اهمیت نتایج پژوهش پیش‌رو از آن جهت است که بلوچل و همکاران (۱۹۹۱) و مارچاند و کلینگویت (۲۰۰۱)، در مطالعات خود به معنی‌داربودن رابطه خطی عمق برف با ارتفاع دست یافته‌اند؛ اما برخی محققان از جمله شابان و همکاران (۲۰۰۴) این رابطه را کلاً مستقیم و برخی دیگر مانند اریکسون و همکاران (۲۰۰۵) آن را معکوس دانسته‌اند. با توجه به نتایج به‌دست‌آمده از مطالعات حوضه مورد نظر در مقایسه با نتایج سایر محققان از جمله بلوچل و همکاران (۱۹۹۱) و شابان و همکاران (۲۰۰۴) که رابطه بین عمق برف و زاویه شیب را معکوس ذکر کرده‌اند و همچنین نتایج به‌دست‌آمده از پژوهش‌های اریکسون و همکاران (۲۰۰۵) و مارچاند و کلینگویت (۲۰۰۱) که این رابطه را مستقیم بیان کرده‌اند و همان‌طور که قرایی‌منش و همکاران (۲۰۱۶) در مطالعه خود شیب را در عمق برف مناسب دانسته‌اند، می‌توان به این نتیجه رسید که علت تفاوت بین نتایج محققان مختلف، تأثیر شرایط منطقه‌ای متفاوت در نتایج مربوط به عمق برف بوده است.

نتیجه‌گیری

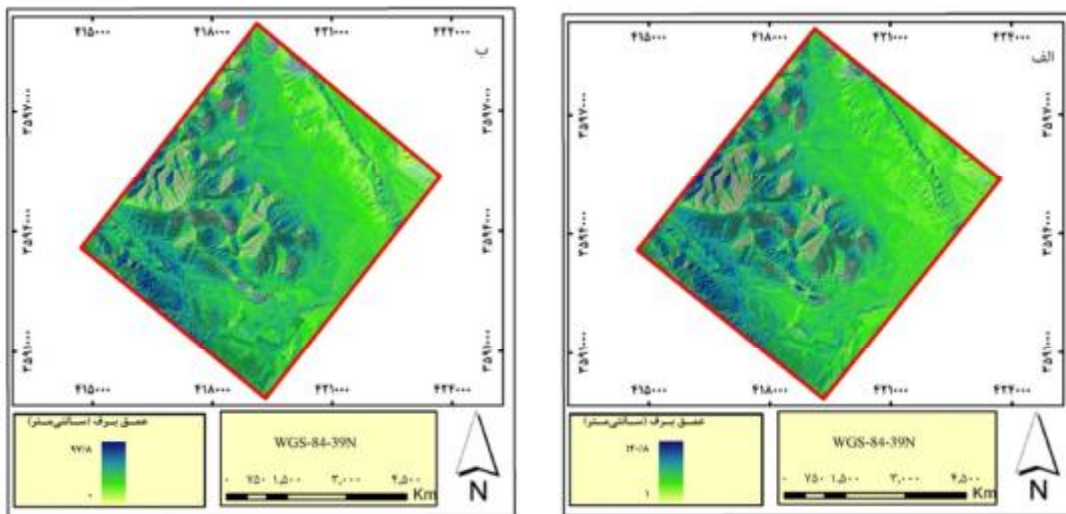
در این پژوهش به‌منظور برآورد توزیع مکانی عمق برف، از روش‌های پارامتریک (رگرسیون خطی چندمتغیره)، ناپارامتریک (رگرسیون مبتنی بر میانگین، رگرسیون مبتنی بر هسته و رگرسیون K- همسایه نزدیک‌تر) و انتخاب ویژگی (استخراج ویژگی PCA، انتخاب ویژگی پیش‌رو و انتخاب ویژگی بهینه‌ساز جمعیت ذره‌ها) استفاده شد. پس از بررسی روش‌های ذکرشده و با توجه به آماره‌های مورد استفاده درجهت ارزیابی مدل‌ها، کمترین میزان مجذور میانگین مربعات خطا مربوط به روش انتخاب ویژگی حریصانه پیش‌رو و بیشترین مجذور میانگین مربعات خطا مربوط به روش رگرسیون خطی چندمتغیره است؛ بنابراین، بی‌شک ویژگی‌های تشریح‌گر نمونه‌ها تأثیر فوق‌العاده‌ای روی دقت روش‌های یادگیری ماشین داشته است و با کاهش و انتخاب پارامترهای مؤثرتر توانسته است عمق برف را با دقت خیلی خوبی برآورد کند.

جدول ۴- نتایج به دست آمده از روش کاهش ویژگی‌ها

MdMRE	MMRE	RMSE	روش یادگیری	مدل
۳۹/۷۱	۴۰/۹۷	۳۷/۵۵	رگرسیون خطی	PCA
۵۳/۶۶	۵۲/۰۱	۴۵/۸۸	رگرسیون مبتنی بر میانگین	
۴۱/۸۷	۴۱/۶۳	۳۹/۰۴	رگرسیون مبتنی بر هسته	
۴۱/۸۳	۴۰/۱۸	۳۷/۳۱	رگرسیون k-همسایه نزدیک‌تر	
۳۷/۹۹	۳۹/۱۷	۲۲/۱۷	رگرسیون خطی	انتخاب ویژگی پیش‌رو
۳۹/۳۱	۴۱/۱۵	۴۴/۸۱	رگرسیون مبتنی بر میانگین	
۳۵/۰۶	۳۶/۴۸	۳۶/۰۴	رگرسیون مبتنی بر هسته	
۳۵/۸۹	۳۷/۰۹	۴۹/۳۵	رگرسیون k-همسایه نزدیک‌تر	
۲۷/۹۳	۳۰/۸۳	۲۲/۱۹	رگرسیون خطی	انتخاب ویژگی ازدحام ذرات
۴۶/۳۰	۴۷/۰۱	۴۳/۱۸	رگرسیون مبتنی بر میانگین	
۳۷/۰۲	۳۷/۱۲	۳۵/۰۲	رگرسیون مبتنی بر هسته	
۳۶/۸۵	۳۵/۰۵	۳۴/۴۷	رگرسیون k-همسایه نزدیک‌تر	



شکل ۲- رابطه رگرسیونی بین عمق برف اندازه‌گیری شده با برآورد شده با استفاده از روش‌های الف- پیش‌رو، ب- ازدحام ذرات



شکل ۳- نقشه پهنه‌بندی عمق برف با استفاده از روش‌های الف) پیش‌رو ب) ازدحام ذرات

10. Bloschl G. Kirnbauer R. and Gutknecht D. 1991. Distributed Snowmelt Simulations in an Alpine Catchment: 1. Model Evaluation on the Basis of Snow Cover Patterns," Water Resources Research. 27(12): 171-179.
11. Elish M. O. Helmy T. and Hussain M. I. 2013. Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation, Math. Probl. Eng.
12. Erickson T. A. Williams M. W. and Winstral A. 2005. Persistence of topographic controls on the spatial distribution of snow in rugged mountain, Colorado, United States. Water Resources Research. 41(4):1-17.
13. Ganjkanlo H. Vafakhah M. Zeinivand H. and Fathzadeh A. 2020. The effect of different sampling schemes on estimation precision of snow water equivalent (SWE) using geo statistics techniques in a semi-arid region of Iran. Geocarto International. 35(16): 1-14
14. Gharaei-Manesh S. Fathzadeh A. and Taghizadeh-Mehrjardi R. 2016. Comparison of artificial neural network and decision tree modles in estimating spatial distribution of snow depth in a semi-arid region of Iran. Cold Regions Science and Technology. 122: 26-35.
15. Haghizadeh A. Keiani A. and Keiani M. 2017. Evaluating the efficiency of geostatistical methods in order to estimate the spatial distribution of snow depth and density in mountainous areas (case study: Gosh Bala watershed of Mashhad). Journal of Hydrogeomorphology. 4(12): 45-66.
16. Kennedy J. and Eberhart R. 1995. Particle Swarm Optimization. pp. 1942-1948.
17. Kirsopp C. Shepperd M. J. and Hart J. 2002. Search heuristics, case-based reasoning and software project effort prediction.
18. Lundquist J. D. and Dettinger M. D. 2005. How snowpack heterogeneity affects diurnal streamflow timing. Water Resour.
19. Marchand W. D. and Killingtveit A. 2001. Analyses of the Relation Between Spatial Snow Distribution and Terrain Characteristics, 58th Estern Snow Conference Ottawa, Ontario, Canada.
20. McKay G. A. and Gray D. M. 1981. The distribution of the snow cover, in: Handbook of Snow, edited by: Gray D. and Hale D. Pergamon Press Canada Ltd. 153-190.
21. McKay M. D. Beckman R. J. Conover W. J. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 1: 239-245.
22. Pandey P. 2013. Analysis of the techniques for software cost estimation, in Advanced

منابع

۱. اصغری سراسکانرود ص. صفری ش. و ملانوری ا. ۱۴۰۰. اندازه‌گیری عمق برف و ارزیابی ارتباط مؤلفه دما با ویژگی‌های برف در حوزه آبخیز ليقوان. نشریه پژوهش‌های حفاظت آب و خاک. ۲۸(۴): ۱۸۷-۲۰۶.
۲. حلبیان ا. ح. و صلحی س. ۱۳۹۹. بررسی ارتباط برف‌پوش (SC) و دمای سطح زمین (LST) با مؤلفه توپوگرافیکی ارتفاع در ارتفاعات البرز مرکزی. پژوهش‌های ژئومورفولوژی کمی. ۹(۲): ۲۴۹-۲۲۹.
۳. حیدری م. هنربخش ا. پژوهش م. و زنگی‌آبادی م. ۱۳۹۷. توسعه مدل بهینه برای استفاده صحیح از منابع زیست‌محیطی در حوزه آبخیز چلگرد. فصلنامه علوم محیطی. ۱۶(۱): ۲۹-۴۶.
۴. طبری ح. معروفی ص. زارع ابیانه ح. امیری چایجان ر. شریفی م. و آخوندعلی ع. ۱۳۸۸. مقایسه روش رگرسیون غیرخطی با روش‌های هوش محاسباتی در برآورد توزیع مکانی آب معادل برف در سراب کارون. علوم و فنون کشاورزی و منابع طبیعی، علوم آب و خاک. ۱۳(۵۰): ۲۹-۴۰.
۵. فدایی کرمانی ا. خانجانی م. ج. و بارانی غ. ع. ۱۳۹۳. کاربرد الگوریتم K- نزدیک‌ترین همسایگی در پایش خشکسالی بر مبنای شاخص بارش استاندارد (SPI) (مطالعه موردی: شهرستان بم). فصلنامه بین‌المللی پژوهش تحلیلی منابع آب و توسعه. ۱: ۱۳۱-۱۳۸.
۶. گنج‌خانلو ح. وفاخواه م. فتح‌زاده ع. و زینی‌وند ح. ۱۳۹۶. مدل‌سازی، تغییرات مکانی آب معادل برف براساس عوامل توپوگرافی و اقلیمی حوضه آبخیز سهرورد استان زنجان. اکوهیدرولوژی. ۴(۳): ۶۸۶-۶۷۳.
7. Alpaydin E. 2014. Introduction to machine learning. MIT press.
8. Balk b. and Elder k. 2000. Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. Water. Resour. Res. 36:13-26.
9. Bavay M. Gr'unewald T. and Lehning M. 2013. Response of snow cover and runoff to climate change in high Alpine catchments of Eastern Switzerland, Adv. Water Resour. 55: 4-16.

- Computing and Communication Technologies (ACCT), Third International Conference on. pp. 16-19.
23. Papatheocharous E. Papadopoulos H. and Andreou A. S. 2014. feature subset selection for software cost modelling and estimation, arXiv Prepr. arXiv1210.1161, 2012.
 - Alpaydin E., Introduction to machine learning.
 24. Poli R. Kennedy J. and Blackwell T. 2007. Particle Swarm Optimization. *Swarm Intelligence*. 1(1): 33-57.
 25. Pomeroy J. W. and Gray D. M. 1995. Snowcover Accumulation, Relocation and Management. National Hydrology Research Institute Science Report No. 7, Environment Canada, Saskatoon.
 26. Puggini L. and McLoone S. 2017. Forward Selection Component Analysis: Algorithms and Applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39(12).
 27. Sarwar S. and Gupta M. 2013. Proposing effort estimation of COCOMO II through perceptron learning rule, *Int. J. Comput.* 70(1): 29-32.
 28. Schweizer J. Jamieson J. B. and Schneebeli M. 2003. Snow avalanche formation, *Rev. Geophys.* 41: 10-16.
 29. Seligman G. 1936. Snow Structure and Ski Fields. International Glaciological Society. Cambridge. UK.
 30. Shaban A. Faour G. Khawlie M. and Abdallah C. 2004. Remote sensing application to estimate the volume of water in the form of snow on Mount Lebanon. *Hydrological Sciences Journal*. 49(4): 643-653.
 31. Sharma A. and O'Neill R. 2002. A nonparametric approach for representing interannual dependence in monthly streamflow sequences. *Water Resour. Res.* 38: 5-15.
 32. Tedesco M. Pulliainen J. Takala M. Hallikainen M. and Pampaloni P. 2004. Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote Sens. Environ.* 90: 76-85.
 33. Tsai Y. Dietz A. Oppelt N. and Kuenzer C. 2019. Remote sensing of snow cover using Spaceborne SAR: A review. *Remote Sensing*. 11(1456): 1-44.
 34. Wang S. Tang J. and Liu H. 2015. Embedded unsupervised feature selection *Proc. Natl. Conf. Artif. Intell.* 1: 470-476.
 35. Yakowitz S. 1993. Nearest-neighbor estimation for null-recurrent Markov time series. *Stoch. Proc. Appl.* 48: 311-318.

