



حل همزمان خوشه‌بندی و تعیین باندهای بهینه داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات

امین علیزاده نایینی*^۱، فرهاد صمدزادگان^۲، سعید نیازمردی^۳

۱. کارشناس ارشد فتوگرامتری دانشگاه تهران

۲. دانشیار گروه نقشه‌برداری، دانشگاه تهران

۳. دانشجوی کارشناسی ارشد سنجش از دور، دانشگاه تهران

تاریخ پذیرش مقاله: ۱۳۹۱/۳/۲۰

تاریخ دریافت مقاله: ۱۳۸۹/۱۱/۱۸

چکیده

امروزه با رشد و توسعه سیستم‌های سنجش از دور و تولید داده‌های فراطیفی با قدرت تفکیک طیفی بالا انتظار می‌رود که شناسایی پدیده‌ها با دقت بیشتری انجام گیرد و دقت خوشه‌بندی کلاس‌های پوشش زمینی مشابه افزایش یابد. با وجود چنین پیشرفتی، اطلاعات طیفی با ابعاد زیاد همچنان مسئله‌ای چالش‌برانگیز در پردازش و تجزیه و تحلیل داده قلمداد می‌شود، به گونه‌ای که با افزایش ابعاد داده، دقت خوشه‌بندی تا یک حد آستانه افزایش و سپس کاهش می‌یابد. بنابراین به منظور انجام خوشه‌بندی مناسب نیاز است تا به طور همزمان باندهای بهینه این تصاویر نیز انتخاب گردند و یا به عبارت دیگر دو مسئله خوشه‌بندی و انتخاب باندهای بهینه با هم در نظر گرفته شوند. در مقاله حاضر، روش جدیدی برای خوشه‌بندی داده‌های فراطیفی به وسیله الگوریتم بهینه‌سازی توده ذرات ارائه شده است که در آن به طور مشترک موقعیت بهینه مراکز خوشه‌ها و تعداد باندهای بهینه تصاویر فراطیفی به دست می‌آید. قابلیت بالای این الگوریتم در جست‌وجوی سراسری فضای مسئله و حل همزمان توابع هدف مختلف باعث شد تا از آن برای خوشه‌بندی داده‌های فراطیفی استفاده گردد. در پژوهش حاضر، پس از بررسی خوشه‌بندی داده‌ای فراطیفی مبتنی بر الگوریتم توده ذرات، روش پیشنهادی خوشه‌بندی بر مبنای حل همزمان خوشه‌بندی و انتخاب باندهای بهینه معرفی می‌گردد. مقایسه نتایج به دست آمده برای خوشه‌بندی داده‌های سنجنده AVIRIS نشان می‌دهد که نتایج حاصل از الگوریتم بهینه‌سازی توده ذرات و همچنین روش پیشنهادی به ترتیب با مقدار کاپای ۷۴/۲۲ و ۷۶/۵۷ از دیگر الگوریتم‌های استفاده شده در این تحقیق بهتر و کارآمدتر است.

کلیدواژه‌ها: بهینه‌سازی مبتنی بر حرکت توده ذرات، خوشه‌بندی، تصاویر فراطیفی، انتخاب باندهای بهینه.

* نویسنده مکاتبه‌کننده: تهران، دانشگاه تهران، پردیس دانشکده‌های فنی، گروه مهندسی نقشه‌برداری، تلفن: ۰۹۱۳۳۲۲۰۴۲۱

۱- مقدمه

امروزه با پیشرفت‌های جدید فناوری سنجنده‌های فراطیفی، دریافت تصویر با صدها باند طیفی با طول موج‌های باریک و پیوسته، امکان‌پذیر شده است. افزایش باندهای طیفی اطلاعات بیشتری را در زمینه پوشش زمینی و مواد مشابه از لحاظ طیفی فراهم آورده و امکان تشخیص و تمایز عناصر مختلف را با وجود شباهت‌های زیاد ممکن ساخته است (Chang, 2007). استخراج و به‌کارگیری این اطلاعات در بسیاری از کاربردهای مطرح - از جمله طبقه‌بندی ابر^۱ (Ambroise et al., 2000) و یا تشخیص تغییرات^۲ (Ghosh et al., 2011, Canty and Nielsen, 2006) - با استفاده از الگوریتم‌های خوشه‌بندی انجام‌پذیر گشته است.

خوشه‌بندی یکی از روش‌های طبقه‌بندی نظارت‌نشده و فرآیندی است خودکار، که در طی آن مجموعه داده‌ای معین به مجموعه‌ای از کلاس‌ها یا خوشه‌ها تقسیم می‌شود. هدف از طبقه‌بندی داده‌ها در چنین فرآیندی، جداسازی آنها به گونه‌ای است که دو داده در یک خوشه تا حد امکان به هم شبیه و دو داده در دو خوشه متفاوت تا حد امکان از یکدیگر متمایز باشند (Xu and Wunsch, 2005).

روش‌های سنتی خوشه‌بندی در داده‌هایی با ابعاد بالا - که از آن جمله می‌توان به داده‌های فراطیفی اشاره کرد - به دلیل پدیدار شدن بهینه‌های محلی^۳ زیاد، با مشکل همگرا شدن به این بهینه‌ها مواجه می‌شوند (Bezdek et al., 2005). یکی دیگر از دلایل عملکرد ضعیف خوشه‌بندی‌های سنتی، حساسیت به مقادیر اولیه و عدم جست‌وجوی سراسری^۴ آنهاست (Bezdek et al., 2005, Bezdek and Hathaway, 1994). برای حل این مشکل، به تازگی الگوریتم‌های تکاملی از قبیل الگوریتم ژنتیک (Bezdek and Hathaway, 1994) الگوریتم بازپخت فلزات^۵ (Selim and Alsultan, 1991) و الگوریتم‌های فراابتکاری مانند الگوریتم کلونی مورچه‌ها (Runkler, 2005)، زنبورها

(Pham et al., 2007) و الگوریتم بهینه‌سازی توده ذرات (PSO) (Runkler and Katz, 2006, Omran et al., 2005) به طور موفقیت‌آمیزی به کار گرفته شده‌اند. طبق مطالعات انجام‌شده، در میان الگوریتم‌های فراابتکاری، عملکرد الگوریتم‌های مبتنی بر هوش جمعی بهتر از دیگران است (Engelbrecht, 2005). اما از بین الگوریتم‌های هوش جمعی نیز الگوریتم بهینه‌سازی توده ذرات برای مسائل با ماهیت پیوسته - نظیر خوشه‌بندی از الگوریتم کلونی زنبور و مورچه‌ها که دارای ماهیت گسسته هستند - نتایج بهتری به دست می‌دهد (Runkler and Katz, 2006). علاوه بر این الگوریتم، بهینه‌سازی توده ذرات ابزاری برای بهینه‌سازی جمعیت مبناست که به دلیل قابلیت بالا در جست‌وجوی سراسری و محلی فضای مسئله، توانایی بالایی در حل مسائل بهینه‌سازی دارد (Kennedy and Eberhart, 1995).

در پژوهش حاضر با تکیه بر الگوریتم بهینه‌سازی توده ذرات، کوششی در مسیر حل مشکل روش‌های خوشه‌بندی سنتی در زمینه داده‌های فراطیفی صورت گرفته است. یکی دیگر از عوامل مؤثر بر دقت روش خوشه‌بندی، تعداد ویژگی‌های مورد استفاده به‌وسیله آن الگوریتم است. به بیان دیگر، دقت خوشه‌بندی یا گروه‌بندی به انتخاب تعداد ویژگی‌های داده‌ها بستگی دارد (Jensen, 2004) و این مسئله در داده‌های فراطیفی با ابعاد بالا اهمیت ویژه‌ای می‌یابد. بنابراین به منظور انجام خوشه‌بندی مناسب برای این گونه تصاویر نیاز است تا علاوه بر بهبود پارامترهای خوشه‌بندی، باندهای بهینه آنها نیز به‌وسیله روش خوشه‌بندی برآورد شود.

1. Cloud Classification
2. Change detection
3. Local optima
4. Global search
5. Simulated annealing
6. Particle swarm optimization

برای خوشه‌بندی برآورد می‌شود (Paoli et al., 2009). لیو و همکاران^۷ نیز روش ترکیبی را برای خوشه‌بندی داده‌های چندطیفی سنجنده MODIS ارائه کرده‌اند روش کار در الگوریتم ایشان به این صورت است که بعد از مجموعه‌ای عملیات پیش‌پردازش نظیر حذف نویز بر روی داده‌ها، خوشه‌بندی فازی که به‌وسیله الگوریتم بهینه‌سازی توده ذرات بهینه شده است برای خوشه‌بندی تصاویر این سنجنده استفاده می‌شود (Liu et al., 2008).

هدف از تحقیق حاضر ارائه روشی جدید بر مبنای الگوریتم بهینه‌سازی توده ذرات است که پارامترهای بهینه خوشه‌بندی و تعداد باندهای بهینه را به طور هم‌زمان به دست آورد. از آنجا که روش خوشه‌بندی مورد استفاده بر مبنای روش‌های خوشه‌بندی جزءبندی است، پارامترهای خوشه‌بندی، همان تعداد خوشه‌ها و موقعیت بهینه همان مراکز خوشه‌ها خواهد بود. از بین دو پارامتر مذکور معمولاً تعداد خوشه‌ها ثابت در نظر گرفته می‌شود، چرا که با مجهول در نظر گرفته شدن تعداد خوشه‌ها عدم قطعیت مسئله بسیار بالا می‌رود و در این حالت به طور کلی ممکن است مسئله همگرا نشود (Paoli et al., 2009). بنابراین در این الگوریتم سعی بر آن است تا ابتدا موقعیت بهینه مراکز خوشه‌ها مشخص شود و سپس تعداد باندهای بهینه تصویر به دست آید.

در مقاله حاضر پس از شرح مختصر الگوریتم بهینه‌سازی توده ذرات، خوشه‌بندی بر مبنای این الگوریتم و روش پیشنهادی ارائه شده است و پس از آن

در سال‌های اخیر تحقیقات گسترده‌ای در خصوص خوشه‌بندی تصاویر فراطیفی انجام پذیرفته است. برای نمونه بیلگین^۱ و همکاران، الگوریتم خوشه‌بندی فازی را با استفاده از تابع شباهت جدیدی برای خوشه‌بندی داده‌های فراطیفی به کار گرفتند. در نوشتار حاضر نیز از تبدیل موجک یک‌بعدی گسسته برای کاهش ابعاد داده‌ها استفاده شده است (Bilgin et al., 2008). الگوریتم خوشه‌بندی فازی وزن‌دار را کوو^۲ و همکاران برای خوشه‌بندی تصاویر فراطیفی پیشنهاد کرده‌اند که در آن مراکز خوشه‌ها با میانگین وزن‌دار خوشه‌ها جایگزین شده است. وزن خوشه‌ها در این الگوریتم با استفاده از الگوریتم استخراج ویژگی و وزن‌دار غیرپارامتریک^۳ محاسبه می‌شود (Kuo et al., 2008). کرافورد و لی^۴ الگوریتم سلسله‌مراتبی دو مرحله‌ای را برای خوشه‌بندی داده‌های فراطیفی پیشنهاد کردند، که در مرحله نخست خوشه‌بندی با ترکیب پیکسل‌های مجاور و در سطح محلی انجام می‌شود و در مرحله دوم با استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی نتایج مرحله قبل خوشه‌بندی می‌شود (Lee and Crawford, 2004). از جمله کارهای انجام‌شده در تصاویر فراطیفی با تکیه بر مفاهیم هوش مصنوعی، در ادامه به موارد مهمی اشاره می‌شود. ژانگ و همکاران^۵ روش خوشه‌بندی دو مرحله‌ای را بر مبنای روش بهینه‌سازی سیستم دفاعی بدن انسان، برای تصاویر فراطیفی و چندطیفی ارائه دادند. نتایج نشان می‌دهد که این الگوریتم در مقایسه با چهار الگوریتم خوشه‌بندی fuzzy c-means, kmeans و Self Organization Map (SOM) عملکرد بالایی در خوشه‌بندی دارد (Zhong et al., 2006). پائولی و همکاران^۶ در سال ۲۰۰۹، الگوریتمی برای خوشه‌بندی تصاویر فراطیفی بر مبنای هوش مصنوعی ارائه کردند که در آن با استفاده از روش خوشه‌بندی آماری به طور هم‌زمان پارامترهای خوشه‌ها، تعداد کلاس‌های داده و تعداد باندهای بهینه

1. Bilgin
2. Kuo
3. Non-parametric weighted Feature Extraction (NWFE)
4. Crawford & Lee
5. Zhong et al
6. Paoli et al
7. Liu

مقیاس‌دهی مؤلفه شناختی و اجتماعی در محاسبه سرعت، r_1 و r_2 مقادیر تصادفی بین صفر و یک هستند. در مسائل بهینه‌سازی مبتنی بر الگوریتم توده ذرات، دو جزء اساسی با نام‌های جمعیت و تابع سازگاری وجود دارد. جمعیت شامل تعدادی ذره است که هر یک پتانسیل کامل حل مسئله را دارند. تابع سازگاری نیز میزان مطلوبیت هر ذره را مشخص می‌کند. لازم به ذکر است که برای هر مسئله، نوع تابع سازگاری و ماهیت ذرات متفاوت است (del Valle et al., 2008).

۲-۲- خوشه‌بندی مبتنی بر الگوریتم توده ذرات
همان‌گونه که توضیحات بخش ۲-۱ نشان می‌دهد، در هر مسئله مبتنی بر الگوریتم توده ذرات دو جزء اساسی وجود دارد که بایستی تعیین گردند. در ادامه، هر یک از این مفاهیم در مسئله خوشه‌بندی توضیح داده خواهند شد. در مفهوم خوشه‌بندی مبتنی بر الگوریتم بهینه‌سازی توده ذرات، هر ذره دربردارنده C ، مرکز خوشه است. بدین ترتیب هر ذره را می‌توان به صورت رابطه (۳) نمایش داد:

رابطه (۳)

$$v_i = (v_{i1}, \dots, v_{ij}, \dots, v_{iC})$$

که v_{ij} ، نشان‌دهنده زامین بردار مرکز خوشه در زامین ذره در خوشه i ، است. بنابراین جمعیت، تعدادی از کاندیداهای خوشه‌بندی را برای داده‌های جاری نمایش می‌دهد. از آنجا که مسئله خوشه‌بندی به‌وسیله الگوریتم بهینه‌سازی توده ذرات با رویکرد حداقل‌سازی حل می‌شود، بنابراین تابع سازگاری هر یک از ذرات طبق رابطه (۴) محاسبه می‌شود.

$$F = 1 - \text{kappa} \quad \text{رابطه (۴)}$$

در رابطه (۴)، kappa مقدار ضریب کاپا است که با توجه به داده‌های واقعیت زمینی، به عنوان دقت برای خوشه‌بندی هر یک از ذرات محاسبه می‌شود.

داده‌ها و پیش‌پردازش آنها و معیارهای ارزیابی‌شان آورده شده است. سپس پارامترهای لازم برای اجرای الگوریتم‌ها و نتایج و بحث در مورد آنها آورده شده است و در پایان نتیجه‌گیری تحقیق آمده شده است.

۲- مواد و روش‌ها

۲-۱- الگوریتم بهینه‌سازی توده ذرات

الگوریتم بهینه‌سازی مبتنی بر حرکت توده ذرات (PSO)، گونه‌ای روش بهینه‌سازی جمعیت مینای الهام گرفته‌شده از رفتار جمعی پرندگان یا ماهی‌هاست که کندی و ابرهات آن را ارائه کرده‌اند (Kennedy and Eberhart, 1995). در الگوریتم بهینه‌سازی توده ذرات، هر ذره پتانسیل کامل حل مسئله را دارد و هر جمعیت از تعدادی ذره تشکیل شده که بیانگر کاندیداهای مختلف حل مسئله است. فضای راه‌حل مسئله به عنوان فضای جست‌وجو به گونه‌ای فرمول‌بندی می‌شود که هر موقعیت در فضای جست‌وجو یک راه‌حل برای مسئله باشد. ذرات برای پیدا کردن بهترین موقعیت در فضا جست‌وجو با هم همکاری می‌کنند تا نهایتاً بر اساس تابع سازگاری مسئله به بهترین جواب برسند. در این الگوریتم هر ذره مطابق با معادلات سرعت و موقعیت اشاره‌شده در روابط (۱) و (۲) در فضای جست‌وجو حرکت می‌کند (Eberhart and Yuhui, 2001).

رابطه (۱)

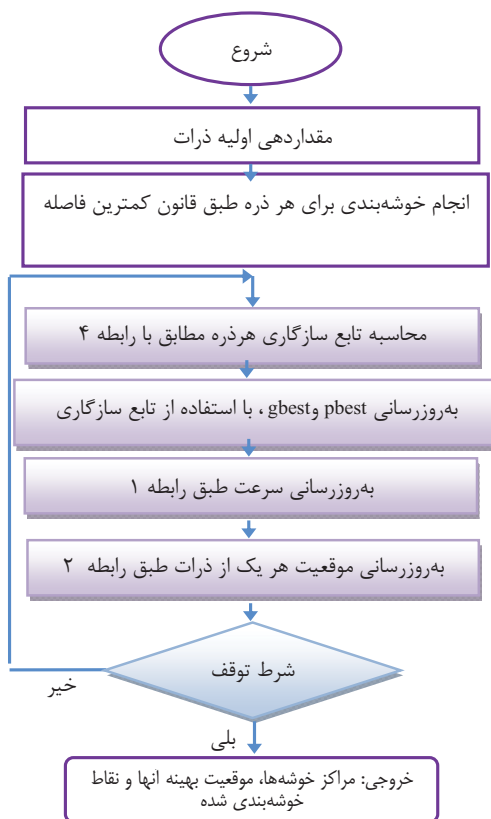
$$v_{id}(t+1) = w v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t))$$

رابطه (۲)

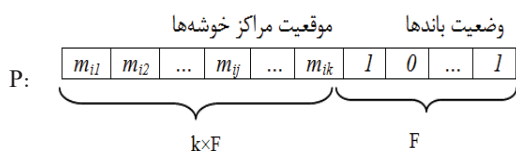
$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1)$$

در رابطه‌های (۱) و (۲) $x_{id}(t)$ و $v_{id}(t)$ به ترتیب موقعیت و سرعت ذره i در زمان t ، $p_{id}(t)$ بهترین موقعیت پیدا شده برای ذره i تا به حال و $p_{gd}(t)$ بهترین موقعیت پیدا شده برای کل جمعیت است و در تمامی متغیرهای تعریف‌شده در روابط مذکور، d نشان‌دهنده ابعاد داده است. W نیز وزن اینرسی مقیاس‌دهنده سرعت مرحله قبل، c_1 و c_2 دو ضریب شتاب برای

حل هم‌زمان خوشه‌بندی و تعیین باندهای بهینه داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات



شکل ۱. نحوه عملکرد خوشه‌بندی PSO



شکل ۲. نمایش یک ذره در حل هم‌زمان خوشه‌بندی و انتخاب باندهای بهینه

اگر k مرکز خوشه وجود داشته باشد، تعداد متغیرهای هر ذره برابر $(k+1) \times d$ خواهد بود، چون هر ذره شامل k مرکز خوشه است که هر یک از آنها خود دارای d بعد هستند. لازم به ذکر است که d متغیر نهایی هر ذره، دارای مقادیر دودویی ۰ یا ۱، بیانگر تعداد باندهای انتخاب شده یا نشده است. به این ترتیب اگر مقدار متغیر برابر صفر باشد، آن بعد در نظر گرفته نمی‌شود و اگر مقدار متغیر یک باشد آن بعد برای آن ذره در نظر گرفته می‌شود. در خصوص جزء

۲-۳- روش پیشنهادی (حل هم‌زمان خوشه‌بندی و انتخاب باندهای بهینه)

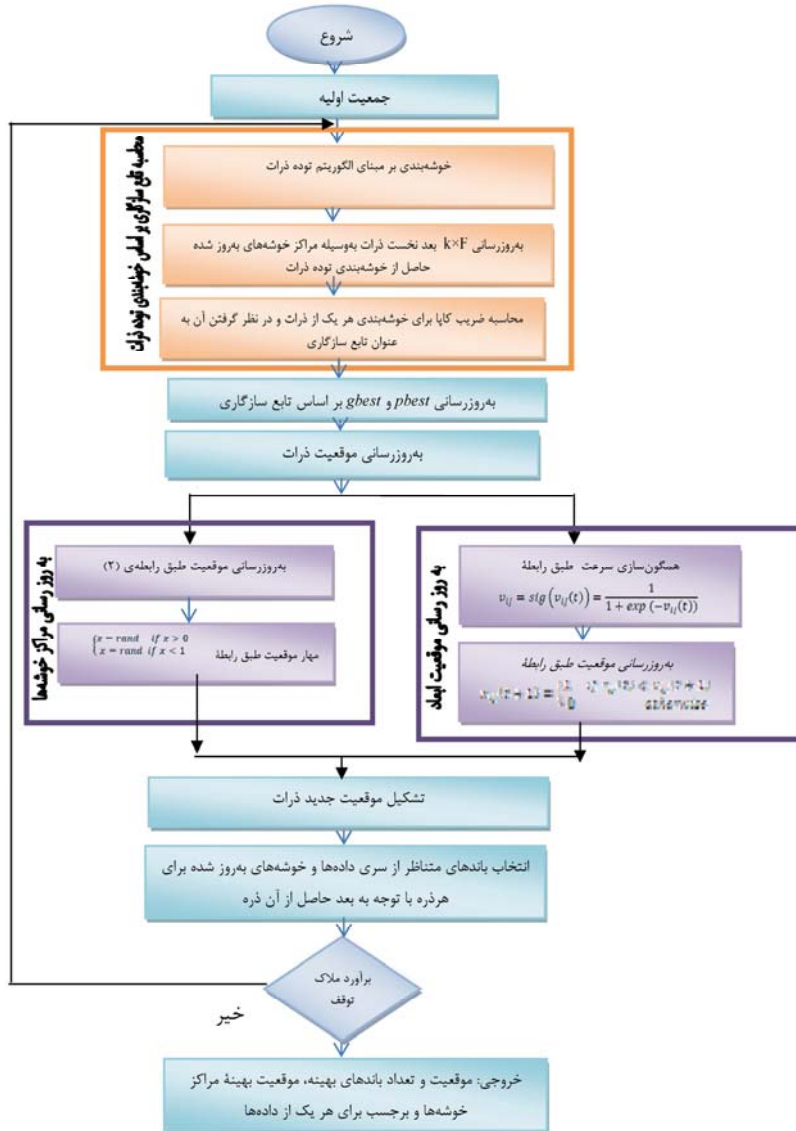
در این روش به‌طور هم‌زمان به حل مسئله خوشه‌بندی و نیز انتخاب باندهای بهینه در تصاویر فراطیفی پرداخته می‌شود. به این منظور از الگوریتم بهینه‌سازی توده ذرات استفاده می‌شود. همان‌طور که قبلاً نیز ذکر شد، در هر مسئله بهینه‌سازی ارائه‌شده در ساختار الگوریتم توده ذرات بایستی دو جزء اساسی موقعیت ذرات و تابع سازگاری تعریف گردد. از آنجا که مسئله خوشه‌بندی مورد نظر در اینجا شامل پیدا کردن موقعیت بهینه مراکز خوشه‌ها و تعداد بهینه باندهای تصویر است، بنابراین هر ذره برداری است که به‌وسیله همه این متغیرها کدگذاری می‌شود. نمایشی از موقعیت یکی از این ذرات در شکل ۲ آورده شده است.

یعنی $w(1-\text{kappa})$ نشان‌دهنده دقت خوشه‌بندی داده‌ها با توجه به زیرمجموعه باندهای انتخاب‌شده است و بخش دوم رابطه زیرمجموعه باندهای انتخاب‌شده به‌وسیله هر ذره را نشان می‌دهد؛ که به ترتیب به کمک پارامتر w و $(1-w)$ وزن‌دار می‌گردند. F در اینجا تعداد کل باندها را نشان می‌دهد. نحوه عملکرد این الگوریتم در شکل ۳ آورده شده است.

اساسی دیگری که تابع سازگاری است در این روش تابع مذکور به صورت رابطه (۵) تعریف می‌شود:

$$F = w(1 - \text{kappa}) + (1 - w) \frac{\sum_{i=k \times F + 1}^{(k+1) \times d} p(l, i)}{F} \quad (5)$$

همان‌طور که در رابطه (۵) نشان داده شده، تابع سازگاری شامل دو قسمت است: بخش یکم این رابطه

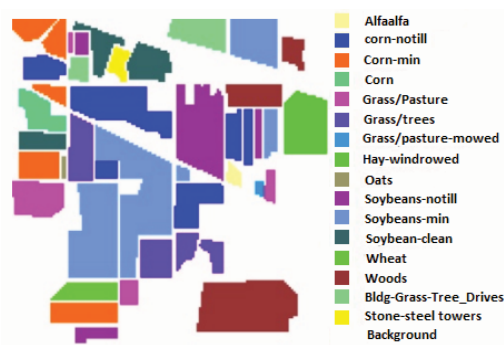


شکل ۳. حل هم‌زمان خوشه‌بندی و انتخاب باندهای بهینه تصویر فراطیفی به‌وسیله الگوریتم توده ذرات

حل هم‌زمان خوشه‌بندی و تعیین باندهای بهینه داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات



(ب)



(الف)

شکل ۴. الف) نقشه واقعیت زمینی ب) ترکیب رنگی منطقه

مورد مطالعه و همچنین نقشه واقعیت زمینی منطقه داده شده است (۱۹۹۲).

این داده دارای ۲۲۰ باند طیفی با پهنای باند ۱۰ نانومتر است و در محدوده طیفی ۰/۴ تا ۲/۴۵ میکرومتر تصویربرداری شده است. ابتدا ۲۰ باند در محدوده جذبی آب و ۱۵ باند نويزدار از مجموع باندهای این داده حذف شده است و ۱۸۵ باند باقی‌مانده در الگوریتم‌های مورد بحث مورد استفاده قرار گرفت (Mojaradi et al., 2008).

از بین کلاس‌های ممکن، پنج کلاس اشاره‌شده در جدول ۱ استفاده می‌شود. دلیل انتخاب این پنج کلاس توزیع نسبتاً مناسب طیفی و فضایی پیکسل‌های مربوط به این کلاس‌ها بوده است (Jia and Richard, 2007).

جدول ۱. اندازه خوشه‌های داده‌ها برای پنج کلاس

نام کلاس	تعداد نمونه
۱ ذرات کاشته‌شده در زمین بدون شخم	۱۴۳۴
۲ درخت‌زار	۷۴۷
۳ یونجه دروشده و خشک‌شده	۴۸۹
۴ لوبیای کاشته‌شده در زمین بدون شخم	۹۶۸
۵ جنگل	۱۲۹۴

۲-۴- داده‌های فراطیفی مورد استفاده

به‌منظور ارزیابی روش پیشنهادی، قطعه تصویری که از منطقه شمال غربی ایندیانا به‌وسیله سنجنده AVIRIS با قدرت تفکیک بیست متر در سال ۱۹۲۲ دریافت شده است، مورد استفاده قرار گرفت. این داده دارای ۱۶ کلاس پوشش زمینی است. از آنجا که کلاس‌های واقع در این قطعه شباهت طیفی نسبتاً بالایی دارند، این داده‌ها برای پژوهشگران به عنوان داده‌هایی چالش برانگیز و مرجع به‌منظور ارزیابی الگوریتم‌های پیشنهادی در زمینه طبقه‌بندی، خوشه‌بندی و کاهش ابعاد فضای اطلاعات مطرح است (Camps-Valls et al., 2007). از آنجا که داده‌ها در ابتدای تابستان برداشت شده‌اند، اکثر محصولات به بیشترین رشد و پوشش زمینی‌شان نرسیده‌اند، بنابراین انرژی اندازه‌گیری‌شده در این داده‌ها فراطیفی هم حاصل از محصولات است و هم از خروجی خاک، رطوبت خاک و علف‌های هرز به دست می‌آید (Mojaradi et al., 2008). در چنین شرایطی طبقه‌بندی به موضوعی چالش برانگیز بدل می‌شود. در شکل ۴ ترکیب رنگی کاذبی از باندهای شماره ۲ (قرمز)، ۱۶۹ (سبز) و ۲۱۱ (آبی) از تصویر

۲-۵- مرحله پیش پردازش

در این تحقیق قبل از پیاده سازی مدل های خوشه بندی، همسان سازی^۱ داده ها انجام شده است. مزیت اصلی این کار در آن است که از تأثیر داده ها در باندهای با محدوده عددی بالاتر بر روی داده ها با محدوده عددی پایین تر جلوگیری می شود. مزیت دیگر آن جلوگیری از پیچیده شدن محاسبات به دلیل تنظیم پارامترهای الگوریتم توده ذرات است. برای مثال می توان به پارامترهای مربوط به محدودیت سرعت و مکان اشاره کرد که در صورت همسان سازی داده ها به راحتی می توان از خروج موقعیت مراکز خوشه ها از محدوده مورد بررسی جلوگیری کرد.

همچنین برای بررسی نحوه عملکرد روش پیشنهادی در استخراج باندهای فراطیفی، الگوریتم های خوشه بندی kmeans، FCM و توده ذرات بعد از به کارگیری مرحله کاهش ابعاد بر اساس ده مؤلفه اول اصلی تولید شده به وسیله آنالیز مؤلفه های اصلی نیز به کار گرفته شدند. دلیل انتخاب این ده مؤلفه این است که با توجه به مقدار ویژه های اندازه گیری شده در محاسبات عناصر اصلی، این ده مؤلفه اصلی دارای ۹۹/۸۱ درصد از محتوای اطلاعاتی داده اصلی هستند.

۲-۶- معیارهای ارزیابی خوشه بندی

شاخص های ارزیابی خوشه بندی را می توان بر اساس سه معیار متفاوت داخلی^۲، نسبی^۳ و خارجی^۴ تعریف کرد. شاخص های داخلی، برازش بین ساختار آشکار شده بین خوشه بندی و داده را ارزیابی می کنند. شاخص های نسبی، ساختارهای چندگانه را با هم مقایسه می کنند (برای مثال خوشه بندی تولید شده به وسیله الگوریتم های مختلف) و تصمیم می گیرند که کدام یک دارای عملکرد بهتری هستند. شاخص های خارجی، عملکرد خوشه بندی را به وسیله تناظریابی ساختار خوشه بندی با یک سری اطلاعات از پیش تعیین شده به نام برجسب های کلاسی صحیح اندازه گیری می کنند (که اغلب به عنوان حقیقت زمینی^۵) در نظر گرفته می شود (jain, 2010, Jain et al., 1999). از آنجا که در

سنجش از دور، هدف از خوشه بندی رسیدن به کلاس هایی با ویژگی های تعیین شده است، استفاده از داده های واقعیت زمینی امری ناگزیری است.

برای ارزیابی نتایج خوشه بندی در این تحقیق، از شاخص های ضریب کاپا (k) و دقت کلی (OA) استفاده شده است. این دو شاخص که از شاخص های پرکاربرد در ارزیابی های خوشه بندی و طبقه بندی تصاویر سنجش از دور هستند (TSO and Mather, 2009) در گروه شاخص های خارجی ارزیابی دقت قابل طبقه بندی هستند (Congalton and Green, 2009). هر دوی این شاخص ها با استفاده از ماتریس خطا محاسبه می شوند اما ضریب کاپا در مقایسه با معیار دقت کلی برآورد کننده بهتری به شمار می آید، به این دلیل که ضریب کاپا بیشتر از معیار دقت کلی، تابع دقت خوشه بندی هر یک از کلاس هاست (Carletta, 1996). در این مقاله نیز برای هر یک از کلاس ها شاخص صحت^۶ برای بررسی استفاده شده است.

$$\hat{k} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad \text{رابطه (۶)}$$

$$k_i = \frac{N \times x_{ii} - x_{i+} \times x_{+i}}{N \times x_{i+} - x_{i+} \times x_{+i}} \quad \text{رابطه (۷)}$$

در رابطه (۶) \hat{k} ، ضریب کاپا و در رابطه (۷) k_i ، شاخص صحت برای هر یک از کلاس هاست. در این روابط همچنین r شماره سطر یا ستون در ماتریس خطا، x_{ii} عضو (i,i) از ماتریس خطا، x_{i+} و x_{+i} به ترتیب مجموع اعضای غیر قطر اصلی سطر i و ستون i و N مجموع تمام مشاهدات است (Congalton and Green, 2009).

1. Normalization
2. Internal
3. Relative
4. External
5. Ground truth
6. Khat index

۳- بحث و نتیجه‌گیری

۳-۱- تنظیم پارامترها

در مطالعه حاضر در خصوص پارامترهای مورد استفاده به‌وسیله الگوریتم توده ذرات، از نتایج تحقیقی استفاده شد که در آن نقش پارامترهای مختلف این الگوریتم در خوشه‌بندی مورد بررسی قرار گرفته است (Omran, 2004). لازم به ذکر است که پژوهشگران مختلف و متعددی نتایج این الگوریتم را مورد استفاده قرار داده‌اند (Samadzadegan and Naeini, 2011, Paoli et al., 2009). در مطالعه حاضر در نهایت از پارامترهایی که در جدول ۲ آمده‌اند برای روش‌های خوشه‌بندی مبتنی بر الگوریتم توده ذرات استفاده شده است.

۳-۲- ارزیابی نتایج

برای ارزیابی روش پیشنهادی، از ضریب کاپا، شاخص صحت و معیار دقت کلی استفاده شد. در شکل‌های ۵ و ۶ به ترتیب مقدار ضریب کاپا و شاخص صحت برای روش‌های مختلف نشان داده شده است. همان‌طور که در شکل ۵ مشخص است، روش پیشنهادی (Simultaneous)، با مقدار ضریب کاپای ۷۶/۵۷ و معیار صحت کلی ۸۱/۸۱ دارای عملکرد بسیار بهتری در مقایسه با روش‌های kmeans-PCA، FCM-PCA، PSO-PCA و PSO است. از آنجا که در الگوریتم پیشنهادی، به‌طور هم‌زمان با خوشه‌بندی و با توجه به معیار ارزیابی آن (یعنی ضریب کاپا)، باندهای بهینه نیز

مشخص می‌شود، نخست خوشه‌بندی داده‌های مورد نظر به دقت بالاتری می‌رسد و دوم با توجه به عملکرد خوشه‌بندی، تعداد باندهای بهینه به‌منظور استخراج داده‌ها مشخص می‌شود (جدول ۳). در نتیجه در این حالت ضمن اینکه دقت خوشه‌بندی بیشتر خواهد شد، به دلیل انتخاب باندهای بهینه، از حجم داده‌های فراطیفی نیز کاسته می‌شود. بعد از روش پیشنهادی، روش PSO، با مقدار ضریب کاپای ۷۴/۲۲ دقت بالاتری در مقایسه با روش‌های دیگر دارد، و بعد از آن روش‌های خوشه‌بندی PSO-PCA و kmeans-PCA دارای عملکرد مناسبی هستند و نهایتاً الگوریتم FCM-PCA با میزان ضریب کاپای ۶۵/۹۴ در مقایسه با روش‌های دیگر دارای دقت پایین‌تری است.

دلیل عملکرد نسبتاً بهتر الگوریتم خوشه‌بندی PSO در مقایسه با PSO-PCA آن است که خوشه‌بندی PSO به تمام فضای داده‌ها و جست‌وجو در آنها دسترسی دارد، در حالی که فضای داده‌های الگوریتم PSO-PCA محدودتر است. اما از آنجا که فضای داده‌های روش PSO-PCA بر اساس آنالیز مؤلفه‌های اصلی قرار دارد، باندهای انتخاب‌شده دارای بیشترین اطلاعات تصویر هستند و به همین دلیل دقت این روش تفاوت چندانی با روش PSO ندارد. از آنجا که داده‌های مورد بررسی هم‌پوشانی ندارند، طبیعی است که الگوریتم فازی این مطالعه (FCM-PCA) عملکرد پایین‌تری در مقایسه با الگوریتم‌های سخت این تحقیق یعنی kmeans+PCA، PSO و PSO+PCA دارد.

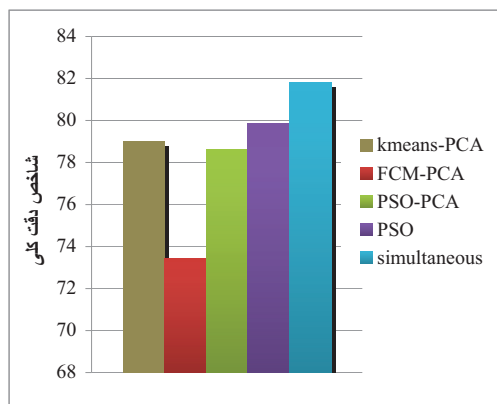
جدول ۲. پارامترهای استفاده‌شده در روش‌های مختلف

پارامترهای استفاده‌شده در روش‌های مبتنی بر الگوریتم توده ذرات	
۱۰۰	تعداد تکرار
۲۰	تعداد ذرات
۰/۷۲	W
۱	c _۱
۱	c _۲
پارامترهای استفاده‌شده در روش kmeans	
۱۰۰	تعداد تکرار
پارامترهای استفاده‌شده در روش FCM	
۱۰۰	تعداد تکرار
۲	پارامتر فازی‌ساز

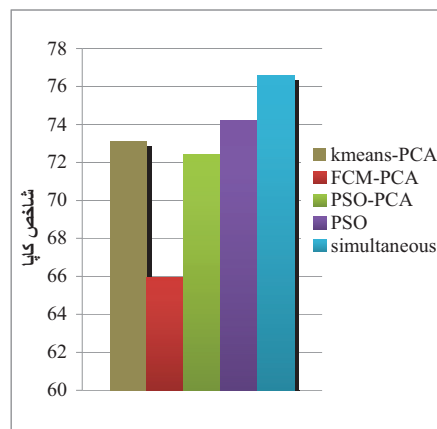
امین علیزاده نایینی و همکاران

جدول ۳. دقت طبقه‌بندی (K) و (OA) و باندهای انتخاب‌شده بر مبنای سه الگوریتم فراابتکاری پیشنهادی

	تعداد باندها	کاپا	صحت کلی
Kmeans+PCA	۱۰	۷۳/۱۳	۷۹/۰۳
FCM+PCA	۱۰	۶۵/۹۴	۷۳/۴۲
PSO+PCA	۱۰	۷۲/۴۲	۷۸/۶۳
PSO	۱۸۵	۷۴/۲۲	۷۹/۸۵
Simultaneous	۸۱	۷۶/۵۷	۸۱/۸۱

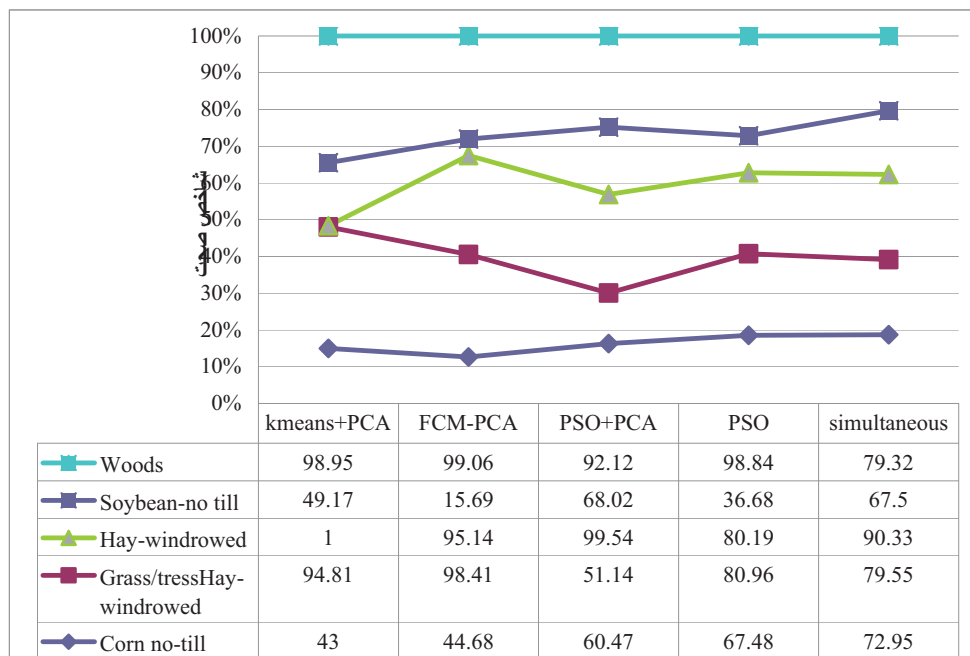


(ب)



(الف)

شکل ۵. (الف) مقایسه مقادیر کاپای روش پیشنهادی دیگر روش‌ها، (ب) مقایسه دقت کلی روش پیشنهادی با دیگر روش‌ها

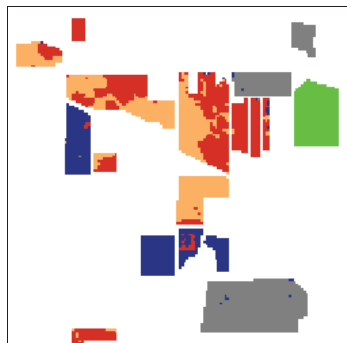


شکل ۶. مقادیر شاخص صحت برای خوشه‌بندی به‌وسیله هر یک از روش‌های مورد بررسی در این تحقیق

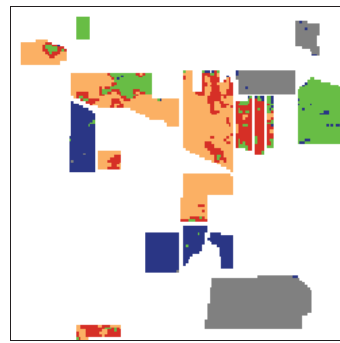
حل هم‌زمان خوشه‌بندی و تعیین باندهای بهینه داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات

استخراج روش‌های مختلف با هم برابر است. نگاشت‌های طبقه‌بندی‌شده حاصل از روش‌های simultaneous, kmeans+PCA, PSO.FCM+PCA و PSO+PCA به ترتیب در شکل‌های ۷-الف تا ۷-ر آورده شده است.

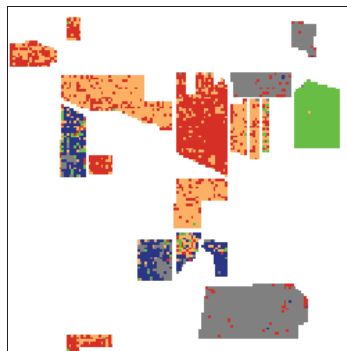
به منظور بررسی بصری، نتایج حاصل از خوشه‌بندی روش‌های مختلف در شکل ۷ آورده شده است. با توجه به شکل حاصل از خوشه‌بندی و میزان شاخص صحت (شکل ۶) مشخص است که روش خوشه‌بندی پیشنهادی به خوبی قادر به تشخیص کلاس‌های مختلف بوده است؛ به گونه‌ای که دقت



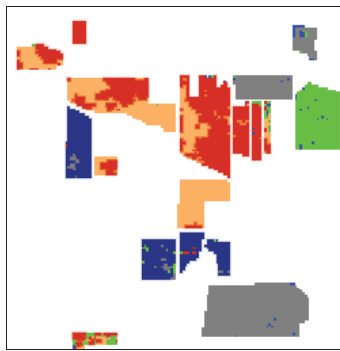
(ب)



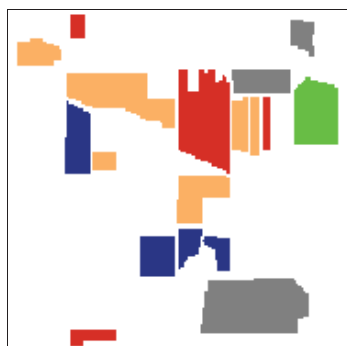
(الف)



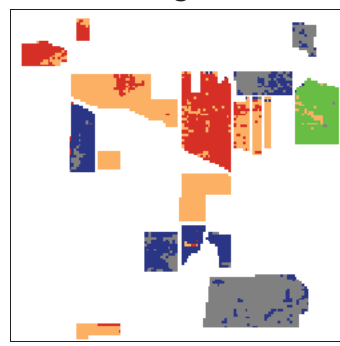
(د)



(ج)



(ز)



(ر)

شکل ۷. نتایج حاصل روی داده‌های فراطیفی دریافت شده به وسیله سنجنده AVIRIS نگاشت‌های طبقه‌بندی حاصل از روش‌های (الف) FCM، (ب) kmeans، (ج) PSO، (د) PSO+PCA، (ر) simultaneous و (ز) نقشه واقعیت زمینی مربوط به آن منطقه

۵- منابع

- AVIRIS image Indian Pine Test Site, 1992, <http://cobweb.ecn.purdue.edu/~biehl/MultiSpectral/hyperspectral.html>.
- Ambrose, C., Sèze, G., Badran, F. & Thiria, S., 2000, **Hierarchical Clustering of Self-organizing Maps for Cloud Classification**, *Neurocomputing*, 30, 47-52.
- Bezdek, J. C. & Hathaway, R. J., 1994, **Optimization of Fuzzy Clustering Criteria Using Genetic Algorithms**, In: *Evolutionary Computation, IEEE World Congress on Computational Intelligence, Proceedings of the First IEEE Conference on, 27-29 Jun, 589-594*.
- Bezdek, J. C., Keller, J. M., Krishnapuram, R. & Pal, N. R. (eds.), 2005, **Fuzzy Models and Algorithms for Pattern Recognition and Image Processing**, Springer.
- Bilgin, G., Erturk, S. & Yildirim, T., 2008, **Unsupervised Classification of Hyperspectral-Image Data Using Fuzzy Approaches That Spatially Exploit Membership Relations**, *Geoscience and Remote Sensing Letters, IEEE*, 5, 673-677.
- Camps-Valls, G., Bandos Marsheva, T. & Zhou, D., 2007, **Semi-Supervised Graph-Based Hyperspectral Image Classification**, *Geoscience and Remote Sensing, IEEE Transactions on*, 45, 3044-3054.
- Canty, M. J. & Nielsen, A. A., 2006, **Visualization and Unsupervised Classification of Changes in Multispectral Satellite Imagery**, *International Journal of Remote Sensing*, 27, 3961-3975.

الگوریتم پیشنهادی این تحقیق در مقایسه با الگوریتم پیشنهادی پائولی و همکاران دارای محاسنی است؛ نخست اینکه الگوریتم پیشنهاد شده در تحقیق از روش‌های آماری برای تخصیص پیکسل‌ها به کلاس‌ها استفاده نمی‌کنند، و در نظر گرفتن توزیع نرمال برای داده‌های فراطیفی - آن گونه که در روش پیشنهادی پائولی و همکاران استفاده شده است - نتایج ضعیفی را به همراه خواهد داشت (Manolakis and Marden, 2002). دیگر اینکه در اینجا برخلاف روش پائولی - که به منظور نشان دادن عملکرد بهتر الگوریتمش داده‌های اشتباهی را به عنوان واقعیت زمینی در نظر گرفته - از داده‌های واقعیت زمینی اصلی برای ارزیابی نتایج استفاده شده است.

۴- نتیجه‌گیری

در مقاله حاضر روش جدیدی برای خوشه‌بندی داده‌های فراطیفی ارائه شده است. با استفاده از این روش می‌توان به طور هم‌زمان مسئله خوشه‌بندی و انتخاب باندهای بهینه را حل کرد. همچنین در این روش از میان روش‌های مختلف خوشه‌بندی مانند پارامتریک و سلسله‌مراتبی، روش خوشه‌بندی جزءبندی انتخاب شده، که حجم محاسباتی آن کمتر از روش‌های مذکور است، بنابراین برای داده‌های فراطیفی با حجم بالا بسیار مناسب به نظر می‌رسد.

نتایج نشان دادند که روش خوشه‌بندی پیشنهادی، علاوه بر آنکه داده‌ها را با دقتی بهتر از چهار روش PSO، PSO+PCA، FCM+PCA، kmeans+PCA و بخش‌بندی می‌کند، حجم داده‌ها را نیز متناسب با روش خوشه‌بندی و یا به عبارت دیگر گروه‌بندی داده‌های مورد استفاده کاهش می‌دهد؛ در صورتی که در روش FCM+PCA، kmeans+PCA، PSO+PCA اگر چه ابعاد داده‌ها بر اساس تحلیل مؤلفه‌های اصلی کاهش پیدا می‌کند، اما چون این کاهش متناسب با هدف مورد نظر - یعنی خوشه‌بندی - نیست، نمی‌تواند دقت خوشه‌بندی را بهبود ببخشد.

- Carletta, J., 1996, **Assessing Agreement on Classification Tasks: The Kappa Statistic**, Computational linguistics, 22, 249-254.
- Chang, C., I (ed.), 2007, **Hyperspectral Data Exploitation: Theory and Applications**, John Wiley & Sons, Inc.
- Congalton, R. G. & Green, K., 2009, **Basic Analysis Techniques. Assessing the Accuracy of Remotely Sensed Data : Principles and Practices**, 2 ed.: CRC Press.
- Del Valle, Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J. C. & Harley, R. G., 2008, **Particle Swarm Optimization: Basic Concepts, Variants and Applications in Power Systems**, Evolutionary Computation, IEEE Transactions on, 12, 171-195.
- Eberhart & Yuhui, S., 2001, **Particle Swarm Optimization: Developments, Applications and Resources**, In: Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, 2001 2001. 81-86 vol. 1.
- Engelbrecht, A. P., 2005, **Fundamentals of Computational Swarm Intelligence**, Wiley Chichester, UK.
- Ghosh, A., Mishra, N. S. & Ghosh, S., 2011, **Fuzzy Clustering Algorithms for Unsupervised Change Detection in Remote Sensing Images**. Information Sciences, 181, 699-715.
- Jain, A. K., 2010, **Data Clustering: 50 Years Beyond K-means**, Pattern Recognition Letters, 31, 651-666.
- Jain, A. K., Murty, M. N. & Flynn, P. J., 1999, **Data Clustering: a Review**, ACM computing surveys (CSUR), 31, 264-323.
- Jensen, R., 2004, **Combining Rough and Fuzzy Sets for Feature Selection**. Ph.D. University of Edinburgh.
- Jia, X. & Richards, J. A., 2007, **Hyperspectral Data Representation**, In: CHANG, C., I (ed.) Hyperspectral Data Exploitation: Theory and Applications. John Wiley and sons, inc.
- Kennedy, J. & Eberhart, R., 1995, **Particle Swarm Optimization**, In: IEEE Int. Conf. Neural Netw. (ICNN), Piscataway, NJ: IEEE, 1942-1948.
- Kuo, B. C., Huang, H. Y., Liu, H. & Tseng, S. C., 2008, **A Novel Fuzzy C-Means Method for Hyperspectral Image Classification**, In: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 7-11 July, 1002-1005.
- Lee, S. & Crawford, M. M., 2004, **Hierarchical Clustering Approach for Unsupervised Image Classification of Hyperspectral Data**, In: IGARSS, Anchorage, AK., IEEE, 941-944.
- Liu, H., Pei, T., Zhou, C. & Zhu, A. X., 2008, **Multi-temporal MODIS-data-based PSO-FCM Clustering Applied to Wetland Extraction in the Sanjiang Plain**, In: International Conference on Earth Observation Data Processing and Analysis, Wuhan, China 7285, 72854Z.

- Manolakis, D. & Marden, D., 2002, **Non Gaussian Models for Hyperspectral Algorithm Design and Assessment**, In: IGARSS, IEEE, 1664-1666, vol. 3.
- Mojaradi, B., Emami, H., Varshosaz, M. & Jamali, S., 2008, **A Novel Band Selection Method for Hyperspectral Data Analysis**, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37.
- Omran, M., Engelbrecht, A. P. & Salman, A., 2005, **Particle Swarm Optimization Method for Image Clustering**, International Journal of Pattern Recognition and Artificial Intelligence, 19, 297-322.
- Omran, M. G. H., 2004., **Particle Swarm Optimization Methods for Pattern Recognition and Image Processing**, Ph.D. University of Pretoria.
- Paoli, A., Melgani, F. & Pasolli, E., 2009, **Clustering of Hyperspectral Images Based on Multiobjective Particle Swarm Optimization**, Geoscience and Remote Sensing, IEEE Transactions on, 47, 4175-4188.
- Pham, D., Otri, S., Afify, A., Mahmuddin, M. & Al-Jabbouli, H., 2001, **Data Clustering Using the Bees Algorithm**. In: 40th CIRP international manufacturing systems seminar, 2007.
- Runkler, T. A., 2005, **Ant Colony Optimization of Clustering Models**, International journal of intelligent systems, 20, 1233-1251.
- Runkler, T. A. & Katz, C., 2006, **Fuzzy Clustering by Particle Swarm Optimization**, In: Fuzzy Systems, 2006 IEEE International Conference on, 601-608.
- Samadzadegan, F. & Naeni, A. A., 2011, **Fuzzy Clustering of Hyperspectral Data Based on Particle Swarm Optimization**, In: Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on, 6-9 June, 1-4.
- Selim, S. Z. & Alsultan, K., 1991, **A Simulated Annealing Algorithm for the Clustering Problem**, Pattern Recognition, 24, 1003-1008.
- Tso, B. & Mather, P. M., 2009, **Pattern Recognition Principles**, Classification Methods For Remotely sensed Data. 1 ed.: CRC Press.
- Xu, R. & Wunsch, D., 2005, **Survey of Clustering Algorithms**, Neural Networks, IEEE Transactions on, 16, 645-678.
- Zhong, Y., Zhang, L., Huang, B. & Li, P., 2006, **An Unsupervised Artificial Immune Classifier for Multi/hyperspectral Remote Sensing Imagery**, Geoscience and Remote Sensing, IEEE Transactions on, 44, 420-431.