نشریه علمی مهندسی فناوری اطلاعات مکانی

سال دوازدهم، شماره دوم، تابستان ۱۴۰۳

Vol.12, No.2, Summer 2024

۱ -۲۲

مقاله پژوهشی

بهبود الگوریتم خوشه‌بندی *K-Means* با استفاده از الگوریتم ژنتیک به منظور تحلیل مکانی شناسایی لکه‌های نفتی در تصاویر پلاریمتری *SAR*

مهرداد کاوه[۱]، یاسر ابراهیمیان قاجاری[۲*]

۱- دانشجوی دکتری گروه سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

۲- استادیار گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه صنعتی نوشیروانی بابل

تاریخ دریافت مقاله: ۱۴۰۱/۰۳/۲۶ تاریخ پذیرش مقاله: ۱۴۰۱/۱۲/۱۵

چکیده

وجود لکه‌های نفتی در بستر دریاها و اقیانوس‌ها، یکی از نگرانی‌ها و دغدغه‌های اصلی محققان در زمینه اکوسیستم دریایی می‌باشد. در این تحقیق از روش خوشه‌بندی *K-Means* مبتنی بر الگوریتم ژنتیک (*GA*) جهت شناسایی لکه‌های نفتی در سطح دریا استفاده شده است. هدف اصلی ارائه الگوریتم *K-Means* بهبودیافته با الگوریتم *GA*، ایجاد یک جستجوی هوشمند و نه صرفا تصادفی در انتخاب مراکز دسته‌های اولیه می‌باشد تا الگوریتم به خوشه‌های بهینه مسئله دست پیدا کند. برای این منظور ابتدا الگوریتم‌های کاهش نویز اسپکل و استخراج ویژگی، به منظور پیش‌پردازش تصاویر رادار دهانه مصنوعی (*SAR*) اعمال شده‌اند. سپس مرکز خوشه‌های بهینه، با هدف بیشترین فاصله بیرون خوشه‌ای، توسط الگوریتم *GA* تعیین شده‌اند. در نهایت برای تعیین خوشه‌های نهایی، از الگوریتم *K-Means* با هدف بیشترین شباهت درون خوشه‌ای، استفاده شده است. به منظور ارزیابی روش‌های خوشه‌بندی، از داده واقعیت زمینی رقومی‌شده استفاده شده است. همچنین جهت ارزیابی الگوریتم *K-Means* بهبودیافته با *GA* از الگوریتم‌های بهینه‌سازی ازدحام ذرات (*PSO*)، بهینه‌سازی مبتنی بر جغرافیای زیستی (*BBO*)، الگوریتم کلونی زنبور مصنوعی (*ABC*) و روش خوشه‌بندی *K-Means* استاندارد استفاده شده است. نتایج حاصل از الگوریتم *K-Means* بهبودیافته توسط *GA* دارای صحت بیشتری نسبت به سایر الگوریتم‌ها می‌باشد. ویژگی آنتروپی توانسته است دقت کلی ۸۳٫۲۴ را حاصل کند که در قیاس با سایر ویژگی‌ها از دقت کلی کمتری برخوردار است، اما دارای قطعیت و صحت بالاتری می‌باشد. ویژگی‌های یاماگوچی، فریمن و مولفه *C11*، علارغم اینکه دقت کلی ۹۰ درصدی را حاصل کرده‌اند، اما به ترتیب با خطای نوع دوم برابر با ۱۸، ۱۱ و ۱۲ درصدی، صحت کمتری را نسبت به دو ویژگی دیگر نشان داده‌اند. نتایج حاصل از این تحقیق نشان می‌دهد که رویکرد پیشنهادی یادگیری ماشین در مقایسه با معماری‌های سنتی، عملکرد بسیار خوبی در مجموعه داده‌های خوشه‌بندی دارد.

کلیدواژه‌ها : لکه‌های نفتی، تصاویر پلاریمتری *SAR*، انتخاب ویژگی، الگوریتم ژنتیک و الگوریتم *K-Means*

* نویسنده مکاتبه‌کننده: مازندران، بابل، خیابان شریعتی، دانشگاه صنعتی نوشیروانی بابل.

تلفن: ۰۹۱۱۳۵۴۴۰۹۵

Email: y.ebrahimian@nit.ac.ir

۱- مقدمه

لکه‌های نفتی در بستر دریاها یکی از دغدغه‌های اصلی محققان در زمینه اکوسیستم دریایی می‌باشد [۱]. امروزه با پیشرفت بشر در طراحی و ساخت کشتی‌های پیشرفته، استفاده از کشتی و راه دریایی یکی از پرکاربردترین روش‌های مبادلات و صادرات در جهان شناخته می‌شود. حضور بیش از حد کشتی‌ها در سطح آب دریاها و اقیانوس‌ها خطرات زیست‌محیطی را بیش از پیش می‌کند [۲]. همچنین در زمینه صادرات نفت به سایر کشورها، حمل نفت با استفاده از راه دریایی، به عنوان اصلی‌ترین روش ممکن جهت فروش نفت می‌باشند [۳]. در برخی موارد به دلیل مشکلات اجتناب‌ناپذیری که در حین انتقال نفت به وجود می‌آید، باعث آزادسازی حجم وسیعی از نفت خام در سطح دریاها و اقیانوس‌ها می‌شود. این آزادسازی بر اکوسیستم منطقه تاثیر زیادی گذاشته و زندگی موجودات آبزی را در معرض خطر قرار می‌دهد. از این رو، شناسایی و پایش این عارضه مخرب در بستر دریاها، امری لازم و حیاتی می‌باشد [۴].

امروزه سنجش از دور به عنوان یکی از ابزارهای قدرتمند در زمینه استخراج و شناسایی لکه‌های نفتی استفاده می‌شود. داده‌های رادار دهانه مصنوعی (*SAR*)[1] نیز به عنوان داده‌های پرکاربرد در سنجش از دور و در زمینه‌هایی که با ساختار عوارض در ارتباط هستند، مورد استفاده قرار می‌گیرند [۵]. به منظور آنالیز تصاویر *SAR* در استخراج اطلاعات، همواره به الگوریتم‌ها و روش‌هایی جهت بهره‌برداری و به کارگیری این تصاویر نیاز است. از آنجایی‌که تصاویر *SAR* ذاتا دارای نویز اسپکل بوده و هم‌چنین به دلیل توزیع نامشخص مقادیر تصویر در فضای ویژگی، روش‌های مرسوم در تصاویر اپتیک، توان ارائه خروجی‌های مناسب را از تصاویر *SAR* ندارند. هم‌چنین مقادیر بازپراکنش از لکه نفتی بسیار شبیه به مقادیر بازپراکنش دریایی با سطح بسیار آرام و یا پدیده‌های در سطح دریاها می‌باشند که عموما به آن‌ها همزاد[2] گفته می‌شود [۶]. این پدیده‌ها شامل جریان‌های روان در سطح اقیانوس‌ها و گرداب‌ها[3]می‌باشند. در مطالعات زیادی محققان تلاش کردند تا لکه‌های نفتی و مقادیر بازپراکنش از لکه نفتی را از یکدیگر متمایز کنند [۷ و ۸]. بنابراین تمایز کلاس دریا و لکه‌های نفتی از یکدیگر، از جمله کاربردهای مهم سنجش از دور در زمینه پایش دریاها و اقیانوس‌ها می‌باشد.

یکی از ساده‌ترین روش‌های جداسازی و شناسایی لکه‌های سیاه در سطح تصویر، اعمال حدآستانه بر روی مقادیر تصویر می‌باشد. در تحقیق فیسلا[4]و همکاران (۲۰۰۰) مقدار نصف میانگین سطح مقطع رادار نرمال (*NRCS*)[5]بر روی تصویر *SAR* اعمال شده است [۹]. در حالیکه در تحقیق نیرچیو[6]و همکاران (۲۰۰۵) همین مقدار منهای انحراف معیار به عنوان حدآستانه مورد نظر انتخاب شده است [۱۰]. در دیگر روش‌های حدآستانه از حدآستانه‌های محلی استفاده شده است. سولبرگ[7]و همکاران (۲۰۰۷) حدآستانه محلی را با حرکت پنجره با ابعاد مشخص بر روی تصویر و روش هرم چند مقیاسه و نیز خوشه‌بندی مشخص کرده‌اند [۱۱]. در تحقیق توپیزلس[8]و همکاران (۲۰۰۸) حدآستانه محلی در محدوده قطعات مشخص از تصویر معرفی شده است [۱۲]. هم‌چنین دل‌فریت[9]و همکاران (۲۰۰۰) در تحقیق خود از روش شناسایی لبه بر مبنای هیستوگرام مقادیر مناطقی از تصویر که مشکوک به اشیاء سیاه می‌باشند، بهره گرفته‌اند [۱۳]. کانا[10]و همکاران (۲۰۰۴) از یک الگوریتم فازی به منظور

[1] *Synthetic-Aperture Radar*

[2] *Look-alike*
[3] *Eddies*
[4] *Fiscella*
[5] *Normalized Radar Cross-Section*
[6] *Nirchio*
[7] *Solberg*
[8] *Topouzelis*
[9] *Del Frate*
[10] *Kannaa*

انتخاب حدآستانه استفاده کرده‌اند [۱۴].

در تمامی تحقیقات بالا از روش‌های آماری به منظور یافتن مناطق سیاه تصویر استفاده کرده‌اند. لیو[1] و همکاران (۱۹۹۷) و وو[2] و لیو (۲۰۰۳) از تبدیل موجک به منظور شناسایی لکه‌های نفتی بهره گرفته‌اند [۱۵ و ۱۶]. هم‌چنین درود[3] و مرسیر[4] (۲۰۰۷) و آراجو[5] و همکاران (۲۰۰۴) از ویژگی‌های استخراج شده از تبدیل موجک به منظور شناسایی لکه نفتی استفاده کرده‌اند [۱۷ و ۱۸]. بنلی[6] و گارزلی[7] (۱۹۹۹) از آنالیز بافتی انجام شده توسط بعد فرکتال و از الگوریتم چند مقیاسه آنالیز بافت بعد فرکتال، استفاده کردند [۱۹]. همچنین در تحقیق مورغانی[8] و همکاران (۲۰۰۷)، حالت خاصی از آنالیز بافتی بعد فرکتال ارائه شده است [۲۰]. در تحقیق توپیزلس و همکاران (۲۰۰۸) از آنالیز شبکه عصبی مصنوعی (*ANN*)[9] به منظور شناسایی لکه‌های نفتی در تصاویر با حد تفکیک مکانی بالا، استفاده شده است [۱۲].

همان‌طور که ملاحظه می‌شود، در شناسایی لکه‌های نفتی، روش‌های آماری اتوماتیک، *ANN*، روش‌های مبتنی بر احتمالات، سیستم‌های هوشمند و سیستم‌های مبتنی بر منطق فازی از جمله روش‌هایی بوده‌اند که محققان بیشتر روی آن‌ها تمرکز کرده‌اند [۱۲، ۲۱، ۲۲ و ۲۳]. تنها در برخی تحقیقات مانند اسکرانس[10] و همکاران (۲۰۱۲) و ال ابری[11] و همکاران (۲۰۱۲) از الگوریتم *K-Means* برای خوشه‌بندی لکه‌های نفتی استفاده شده است [۲۴ و ۲۵]. هم‌چنین در بسیاری از تحقیقات مانند شیائو[12] و همکاران (۲۰۱۰)، تاری[13] و هاشمی[14] (۲۰۱۰)، کاوه[15] و همکاران (۲۰۲۰) و ایسلام[16] و همکاران (۲۰۱۸) از ترکیب الگوریتم ژنتیک (*GA*)[17] و الگوریتم‌های موجود در داده‌کاوی (*K-Means*)، برای مسائل مختلف استفاده شده است [۲۶، ۲۷، ۲۸ و ۲۹]. اما هیچ یک از این تحقیقات نام‌برده شده مربوط به شناسایی لکه‌های نفتی در تصاویر *SAR* نمی‌باشند. از این رو یکی از انگیزه‌های اصلی این تحقیق، ارائه روشی نظارت‌نشده با استفاده از الگوریتم خوشه‌بندی *K-Means* با تکیه بر الگوریتم *GA*، جهت شناسایی لکه‌های نفتی می‌باشد. هدف از ارائه الگوریتم پیشنهادی، ایجاد یک منطق روشن و نه صرفا تصادفی در انتخاب مراکز دسته‌های اولیه می‌باشد تا الگوریتم به خوشه‌های بهینه مسئله دست پیدا کند. در ادامه مشارکت‌های اصلی مقاله و همچنین ساختار مقاله بیان می‌شوند.

۱-۱- مشارکت‌های اصلی مقاله

مشارکت‌های اصلی مقاله به شرح زیر می‌باشند:

- ارائه روشی نظارت‌نشده با استفاده از الگوریتم *K-Means* مبتنی بر *GA*، جهت شناسایی لکه‌های نفتی.
- بررسی پتانسیل و عملکرد ویژگی‌های پلاریمتری و ترکیب آن با الگوریتم خوشه‌بندی *K-Means*.
- بهبود عملکرد الگوریتم *K-Means* با استفاده از *GA*، به منظور ایجاد یک جستجوی هوشمند و نه صرفا تصادفی در انتخاب مراکز دسته‌های اولیه، تا الگوریتم به خوشه‌های بهینه مسئله دست پیدا کند. به عبارتی دیگر، مرکز خوشه‌های بهینه، با هدف بیشترین فاصله بیرون خوشه‌ای، توسط *GA* تعیین می‌شوند. سپس برای تعیین خوشه‌های نهایی، از الگوریتم *K-Means* با هدف بیشترین شباهت درون خوشه‌ای، استفاده می‌شود.

[1] *Liu*
[2] *Wu*
[3] *Derrode*
[4] *Mercier*
[5] *Araújo*
[6] *Benelli*
[7] *Garzelli*
[8] *Marghany*
[9] *Artificial Neural Network*
[10] *Skrunes*
[11] *Al Abri*
[12] *Xiao*
[13] *Tari*
[14] *Hashemi*
[15] *Kaveh*
[16] *Islam*
[17] *Genetic Algorithm*

• مقایسه و ارزیابی روش خوشه‌بندی پیشنهادی *GA-K-Means* با الگوریتم‌های بهینه‌سازی ازدحام ذرات (*PSO*)[1]، بهینه‌سازی مبتنی بر جغرافیای زیستی (*BBO*)[2]، کلونی زنبور مصنوعی (*ABC*)[3] و روش *K-Means* استاندارد.

۱-۲- ساختار مقاله

مفاد بخش‌های دیگر این مقاله به شرح زیر می‌باشند؛ بخش دوم، روش انجام تحقیق، منطقه مطالعاتی، آماده‌سازی و پیش‌پردازش داده‌ها و اصول اولیه الگوریتم‌های پیشنهادی را شرح می‌دهد. در بخش سوم نتایج این تحقیق آمده است که شامل مدل‌سازی مسئله، کالیبراسیون پارامترهای الگوریتم‌ها، خروجی الگوریتم‌ها و ارزیابی نتایج می‌باشد. در نهایت، در بخش آخر نیز نتیجه‌گیری و پیشنهادهای این تحقیق ارائه شده است.

۲- روش انجام تحقیق

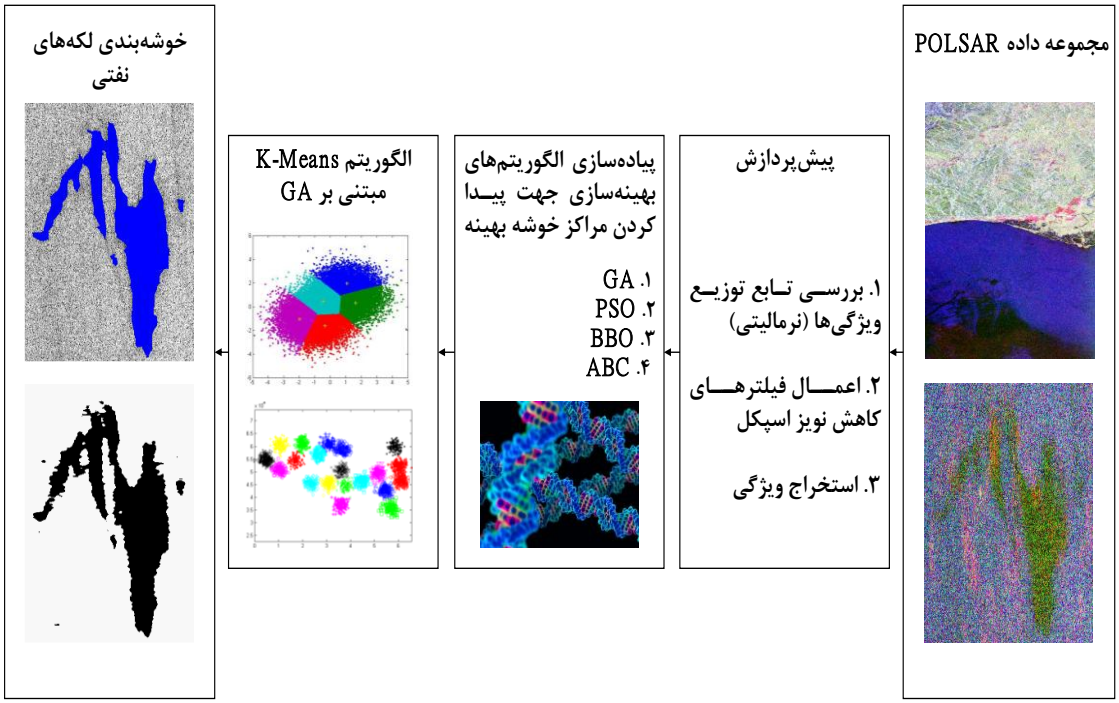
هدف اصلی از این پژوهش، بهبود الگوریتم خوشه‌بندی *K-Means* مبتنی بر الگوریتم *GA* جهت شناسایی لکه‌های نفتی در سطح دریا می‌باشد. در شکل (۱) روند کلی اجرای تحقیق آمده است. همان‌طور که ملاحظه می‌شود، پس از جمع‌آوری داده‌ها، ابتدا پیش‌پردازش بر روی آن‌ها صورت می‌پذیرد. اعمال الگوریتم‌های کاهش نویز اسپکل و استخراج و انتخاب ویژگی، از جمله پیش‌پردازش‌های اصلی تصاویر *SAR* در این تحقیق می‌باشند. در مرحله بعدی برای خوشه‌بندی لکه‌های نفتی از الگوریتم *K-Means* بهبودیافته با *GA* استفاده می‌شود. به این صورت که ابتدا مرکز خوشه‌های بهینه توسط *GA* تعیین شده و سپس از الگوریتم *K-Means* برای تعیین خوشه‌ها استفاده می‌شود. در نهایت نتایج به دست آمده مورد ارزیابی قرار می‌گیرند. در ادامه منطقه مطالعاتی، پیش‌پردازش داده‌های ورودی، مفاهیم اولیه الگوریتم‌های *GA* و *K-Means* بهبود داده شده توسط *GA* به‌طور خلاصه بررسی خواهند شد.

۲-۱- منطقه مورد مطالعه و داده‌ها

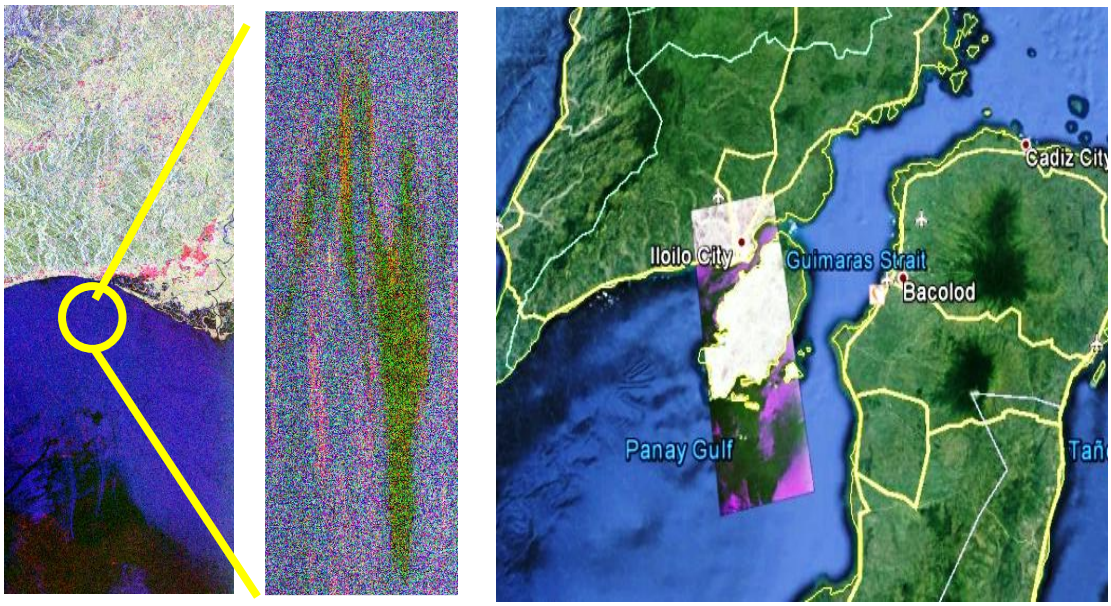
لکه نفتی بررسی شده در این مطالعه مربوط به حادثه‌ای است که کشتی نفتکش از مسیر خلیج پانی[4] به سمت جزیره نگروس[5] در حوزه آبی فیلیپین در حرکت بوده است. غرق شدن کشتی در تاریخ ۱۱ اوت ۲۰۰۶ اتفاق افتاده و لکه نفتی به وجود آمده در تاریخ ۲۵ اوت ۲۰۰۶ به بزرگترین اندازه خود رسیده است. در این تاریخ سنجنده *ALOSPALSAR* تصویری از این منطقه برداشت نموده که در این تحقیق از آن استفاده شده است. این لکه نفتی با مختصات مرکزی *11.12 N 122.30 E* می‌باشد که در شکل (۲) ملاحظه می‌شود.

داده‌های استفاده شده در این تحقیق مربوط به تصاویر سنجنده *ALOSPALSAR* می‌باشند که از سایت *https://vertex.daac.asf.alaska.edu* به صورت آزاد دانلود شده‌اند. این سنجنده قابلیت تصویربرداری به صورت تک کاناله، دو کاناله و چهارکاناله در حالت‌های مختلف با باند *L* را دارد. هم‌چنین از داده‌های واقعیت زمینی نیز به منظور ارزیابی نتایج استفاده شده است. به منظور ارائه داده واقعیت زمین، از دیجیتایز کردن قسمت مشخصی از لکه نفتی در تصویر پائولی *RGB* ساخته شده از تصویر پلاریمتری و تصویر کروگاگر *RGB* بهره گرفته شده است. با توجه به لکه‌های نفتی موردنظر در محدوده، داده واقعیت زمینی از روی تصاویر پائولی و کروگاگر رقومی شده است (شکل(۳)).

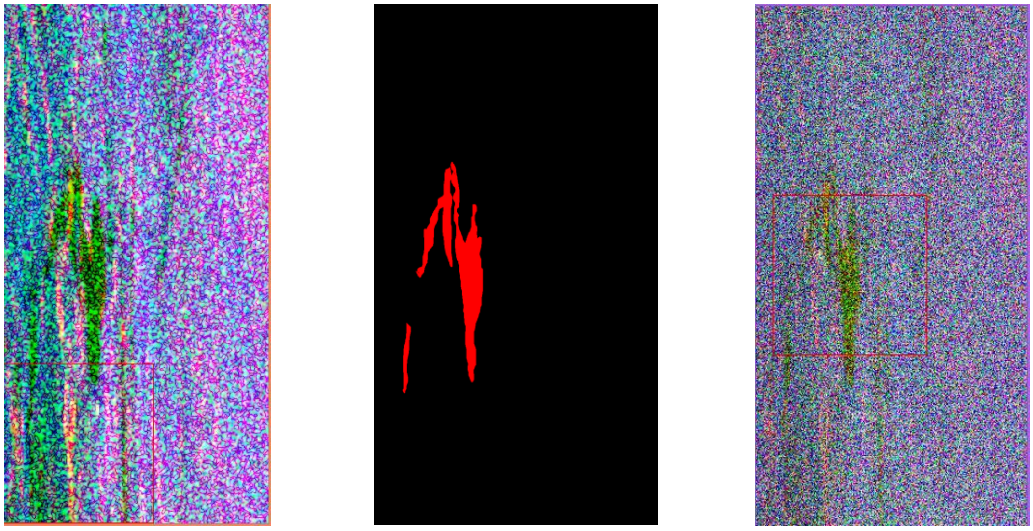
[1] *Particle Swarm Optimization*
[2] *Biogeography-Based Optimization*
[3] *Artificial Bee Colony*
[4] *Panay*
[5] *Negros*



شکل ۱: روند کلی انجام تحقیق



شکل ۲: منطقه مورد مطالعه



(الف) (ب) (ج)

شکل ۳: (الف) تصویر مربوط به ترکیب رنگی پائولی *RGB*، (ب) تصویر موقعیت لکه نفتی رقومی شده از دو تصویر دیگر، (ج) تصویر ترکیب رنگی مولفه تجزیه کروگاگر

۲-۲- پیش‌پردازش داده‌ها

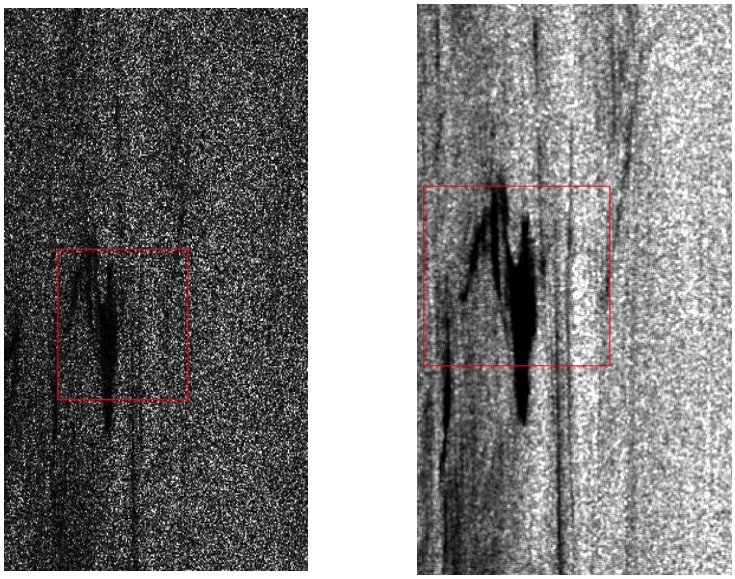
مرحله آماده‌سازی و پیش‌پردازش داده‌ها در خوشه‌بندی، از جمله مهم‌ترین مرحله در فرایند داده‌کاوی می‌باشد. در این تحقیق پیش‌پردازش تصاویر *SAR*، به سه دسته اصلی بررسی تابع توزیع ویژگی‌ها، اعمال الگوریتم‌های کاهش نویز اسپکل و استخراج ویژگی تقسیم می‌شوند. در ابتدا تابع توزیع تمام ویژگی‌ها به عنوان اطمینان حاصل کردن از نوع توزیع نرمال در قسمت پیش‌پردازش بررسی شد. با بررسی تابع توزیع و هیستوگرام‌های ویژگی‌ها، دریافتیم که تصاویر ورودی تا حد قابل قبولی از توزیع نرمال پیروی می‌کنند. بنابراین امکان استفاده از الگوریتم خوشه‌بندی به منظورتشخیص تارگت وجود خواهد داشت. در ادامه سایر پیش‌پردازش‌های انجام‌شده در این تحقیق بررسی می‌شوند.

۲-۲-۱- کاهش نویز اسپکل

تصاویر *SAR* به دلیل ذات تصویر و برهم‌کنش امواج با یکدیگر در طی مسیر ارسال و دریافت سیگنال توسط سنجنده، به صورت ضرب‌شونده، بنا به پدید داپلر همدیگر را تقویت و یا از شدت هم کم می‌کنند. این امر، تصاویر و اطلاعات نهایی که به دهانه سنجنده می‌رسد را تحت تاثیر قرار می‌دهد و اصطلاحا نویز اسپکل در سطح تصویر اتفاق می‌افتد. این نویز بنا بر ادعای برخی مطالعات نباید به صورت یک نویز در تصویر بیان شود. چرا که این عامل مخرب تصویر، در خود ذات و ماهیت تصویر قرار داشته و عاملی جدا ناپذیر در تصویر بیان می‌شود. از این رو مطالعات زیادی الگوریتم‌های کاهش نویز اسپکل را پیشنهاد داده‌اند تا بتوان اثر این عامل بر روی تصویر را کمتر نمود. در این پژوهش، برای کاهش این نویز، از الگوریتم *BoxCar* با ابعاد پنجره ۷ در ۷، بهره گرفته شده است. الگوریتم *BoxCar* از جمله الگوریتم‌های رایج در زمینه کاهش نویز اسپکل می‌باشد که در اکثر مطالعات از آن استفاده شده است [۳۶، ۳۷ و ۳۸].

شکل (۴)، خروجی نهایی تصویر پیش‌پردازش شده با رفع نویز اسپکل را نشان‌می‌دهد. مطابق با شکل (۴)، تصویر اولیه و تصویر پیش‌پردازش شده با اعمال این الگوریتم از یکدیگر متمایز شده‌اند. هم‌چنین، نویز

اسپکل که در تصویر سمت چپ وجود داشته است، تا حدودی پس از اعمال فیلتر از بین رفته و تصویر مورد نظر تارتر به نظر می‌رسد. علت این موضوع نیز عملیات میانگین‌گیری در پنجره با ابعاد تعریف شده است که باعث می‌شود تصویر به اصطلاح تار شود و لبه‌ها تا حدودی از بین بروند.



شکل ۴: تصویر سمت چپ تصویر مولفه *C11* پیش از اعمال فیلتر و تصویر سمت راست تصویر مولفه *C11* پس از اعمال فیلتر

۲-۲-۲- انتخاب ویژگی‌ها

تصاویر پلاریمتری *SAR* در حالت‌ها و پارامترهای مختلفی قابل تجزیه هستند که می‌توانند ویژگی‌های ساختاری سطح را بهتر نشان دهند. علاوه بر ویژگی‌های اصلی این تصاویر، شامل المان‌های اصلی ماتریس *C3*، *T3* و ماتریس پراکنش، ویژگی‌های دیگری نیز قابل استخراج می‌باشند. به طورکلی تفکیک کننده‌های *SAR* و الگوریتم‌های تجزیه هدف از تصاویر پلاریمتری *SAR*، نیز ارائه می‌شوند که هر یک دارای پارامترهایی هستند که بیان کننده ویژگی سطح می‌باشند. جدول (۱) نام هر یک از این دسته ویژگی‌ها و همچنین تعداد هر یک از آن‌ها را بیان می‌کند.

از آنجایی‌که تعداد ویژگی‌های حاصل از تصاویر پلاریمتری بسیار زیاد می‌باشند، می‌بایست تعدادی از این ویژگی‌ها به عنوان ویژگی‌هایی که بیشترین تاثیر را در عملکرد خوشه‌بندی دارند، شناسایی شوند. از این رو در این مقاله ویژگی‌هایی مورد بررسی قرار گرفته‌اند که در تحقیق متکان و همکاران (۲۰۱۳) استخراج شده‌اند [۳۲]. در تحقیق متکان و همکاران، عملکرد ویژگی‌ها شناسایی شده و در این تحقیق نیز از همان ویژگی‌ها به‌عنوان داده‌های ورودی استفاده شده است. این ویژگی‌ها عبارتند از؛ مولفه *C11* از ماتریس کواریانس، ویژگی آنتروپی، المان *ODD* از تجزیه‌های یاماگوچی[1]، فریمن[2] و کروگاگر[3] که بیانگر مکانیزم پراکنش تک سطحی می‌باشند.

[1] *Yamaguchi*
[2] *Freeman*
[3] *krogager*

جدول ۱: نام و تعداد ویژگی‌های حاصل از تصویر پلاریمتری

ویژگی	توصیفات	نماد	تعداد ویژگی
ویژگی‌های اصلی	ماتریس کوواریانس	*[C]*	۹
	ماتریس کوهرنسی	*[T]*	۹
	ماتریس پراکندگی	*[S]*	۴
ویژگی‌های تجزیه	کروگاگر	*[krog]*	۳
	هوینن	*[h]*	۹
	بارنز	*[B]*	۹
	کلود	*[Cl]*	۹
	هولم	*[Hol]*	۹
	وانزیل	*[V]*	۳
	کلود پویتر	*Gamma ،lambda ،alpha ،A ،H ،HA ،anisotropy ، asym ،Delta ،RVI ،(1-H)(1-A) ،H(1-A) ،(1-H)A*	۱۹
	فریمن دوردن	*[Fd]*	۳
	یاماگوچی	*[Y]*	۴
	توزی	*[Toz]*	۴
ویژگی‌های تفکیک	ضرایب همبستگی	*[CC]*	۴
	شدت قطبیدگی	*[Pi]*	۱
	درجه قطبی شدن	*[D]*	۱
	گستردگی	*[S]*	۱

۲-۳- الگوریتم *GA*

الگوریتم ژنتیک ازجمله روش‌های فراابتکاری در مسائل بهینه‌سازی است که اولین بار در سال ۱۹۷۵ به‌وسیله هالند[1] مطرح شده است و در واقع یک شبیه‌سازی مجازی از نظریه تکامل تدریجی داروین می‌باشد [۳۰]. این الگوریتم یک روش مبتنی بر جمعیت می‌باشد که در هر تکرار محاسباتی روی جمعیتی از کروموزوم‌ها عمل کرده و تغییرات تصادفی بر روی آن‌ها از طریق اعمال عملگرهای ادغام و جهش ایجاد می‌کند. سپس جواب‌های مختلف به‌دست آمده بر اساس تابع هدف ارزیابی می‌شوند [۳۹، ۴۰ و ۴۱]. *GA* دارای سه عملگر انتخاب، ادغام و جهش می‌باشد که عبارت‌اند از:

- **انتخاب:** انتخاب عملی است که در آن کروموزوم‌های والد برای تولیدمثل تعیین می‌شوند. روش‌های مختلفی برای انتخاب وجود دارد که از جمله می‌توان به چرخ گردان اشاره کرد [۴۲].
- **ترکیب:** ترکیب عملی است که بر روی دو کروموزوم والد عمل کرده و خصوصیات آن‌ها را از نقطه تقاطع تعویض می‌کند. نقطه تقاطع به‌صورت تصادفی انتخاب می‌شود و می‌تواند تک‌نقطه، دونقطه و یا چندنقطه‌ای باشد [۴۳].
- **جهش:** جهش بخش تصادفی *GA* است که بر روی یکی از کروموزوم‌های جمعیت عمل می‌کند و خصوصیات آن را فقط در نقطه جهش تغییر می‌دهد [۴۴].

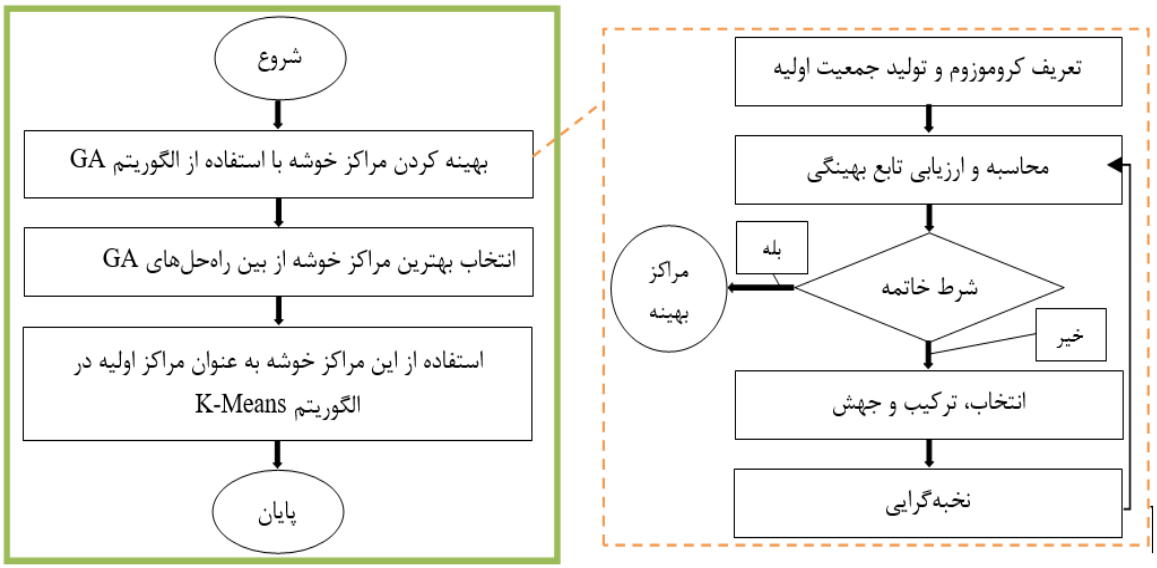
۲-۴- الگوریتم *K-Means* مبتنی بر *GA*

K-Means یک الگوریتم بدون نظارت در داده‌کاوی محسوب می‌شود که به منظور خوشه‌بندی مجموعه

[1] *Holland*

داده‌های بزرگ استفاده می‌شود.

مک‌کوئین الگوریتم *K-Means* را برای اولین بار در سال ۱۹۶۷ ارائه نمود [۳۱]. دراین الگوریتم، ابتدا به تعداد خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شوند. سپس داده‌ها با توجه با میزان شباهت، به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شوند [۲۶]. ایراد اصلی این روش، انتخاب اولیه مراکز خوشه می‌باشد. زیرا کاملا تصادفی انتخاب می‌شوند و بر مبنای هیچ منطق روشنی نمی‌باشند. به عنوان مثال، اگر مراکز خوشه به گونه‌ای بد و دور از مراکز بهینه مسئله انتخاب شوند، الگوریتم حول این نقاط مراکز، خود را بروزرسانی می‌نماید. در نهایت با اتمام شرط خاتمه، خوشه‌های بهینه حاصل نمی‌شوند. در این حالت اصطلاحا گفته می‌شود که الگوریتم در بهینه محلی گرفتار شده است. به همین دلیل در این تحقیق، ابتدا از *GA* برای بهینه کردن مراکز دسته استفاده شده است. سپس این مراکز به عنوان مراکز خوشه‌های اولیه، وارد الگوریتم *K-Means* می‌شوند. در این حالت، با یک اجرای کامل از الگوریتم *K-Means*، می‌توان به احتمال فراوان به خوشه‌های بهینه مسئله دست یافت و یا به آن‌ها نزدیک شد. در واقع الگوریتم *K-Means* به مراکز خوشه اولیه بسیار حساس می‌باشد. هدف اصلی الگوریتم *K-Means* بهبودیافته با *GA*، ایجاد یک منطق روشن و نه صرفا تصادفی در انتخاب مرکز دسته‌های اولیه می‌باشد تا الگوریتم به خوشه‌های بهینه دست پیدا کند و یا حداقل به آن‌ها نزدیک شود. ضمن اینکه مفهوم جستجوی تصادفی نیز در ذات *GA* وجود دارد و از بین نمی‌رود. بنابراین ترکیب جستجوی تصادفی و جستجوی هوشمند، هدف اصلی توسعه الگوریتم پیشنهادی می‌باشد. همان‌طور که در تمامی الگوریتم‌های فراابتکاری، هدف اصلی ایجاد یک موازنه بین دو مفهوم بهره‌برداری (*Exploitation*) و اکتشاف (*Exploration*) می‌باشد، در الگوریتم پیشنهادی نیز، هدف ایجاد یک مدل مبتنی بر جستجوی همسایگی و تصادفی در انتخاب خوشه‌های اولیه می‌باشد. شکل (۵) روند اصلی الگوریتم *K-Means* بهبودیافته با *GA* را نشان می‌دهد.



شکل ۵: روند کلی الگوریتم *K-means* بهبودیافته با *GA*

۳- بحث و نتایج

همان‌طور که ذکر شد، هدف اصلی از این پژوهش، ارائه روشی نظارت‌نشده جهت شناسایی لکه‌های نفتی می‌باشد. در این بخش ابتدا مدلسازی مسئله در قالب *GA* و همچنین تعیین مراکز خوشه بهینه توسط *GA* ارائه می‌شود. سپس نتایج معماری‌های ترکیبی و ارزیابی نتایج آن‌ها ارائه می‌شوند.

۳-۱- پیاده‌سازی الگوریتم *GA*

تعریف یک راه‌حل در قالب کروموزوم یکی از مراحل اصلی در بهینه‌سازی با استفاده از *GA* می‌باشد [۳۳]. در این تحقیق، تعداد ژن در هر کروموزوم، بیان‌گر تعداد ویژگی‌های مرکز خوشه‌های موردنیاز می‌باشد. از آن‌جایی‌که تعداد خوشه‌های مسئله مشخص و برابر با سه می‌باشد، لذا یک کروموزوم از سه مرکز خوشه تشکیل شده است. اما با توجه به اینکه برای هر مرکز خوشه یک فضای ویژگی پنج بعدی وجود دارد، لذا هر سه مرکز خوشه از پنج ویژگی تشکیل شده‌اند. بنابراین یک کروموزوم به‌عنوان یک آرایه پانزده‌تایی متشکل از ویژگی‌های نامزد تعریف شده است. شکل (۶) نمونه‌ای از یک کروموزوم را نشان می‌دهد. مطابق با شکل، پنج ژن اول، ویژگی‌های مربوط به مرکز خوشه اول، پنج ژن وسط، ویژگی‌های مربوط به مرکز خوشه دوم و پنج ژن آخر هم ویژگی‌های مربوط به مرکز خوشه سوم می‌باشند. در این مسئله پنج تصویر وجود دارد که هر کدام دارای ابعاد ۱۳۰۱×۷۰۱ پیکسل می‌باشند و در مجموع ۹۱۲۰۰۱ مقدار ویژگی دارند. بنابراین هر یک از ژن‌های تعریف شده در کروموزوم، از بین ۹۱۲۰۰۱ ویژگی انتخاب می‌شوند. لازم به ذکر است که مقدار هر ژن شماره‌ی پیکسل انتخابی است و تمامی عملگرهای ژنتیک با شماره‌های این پیکسل‌ها سروکار دارند. تنها در تابع هدف مقادیر پیکسل‌ها فراخوانی می‌شوند. برای تولید جمعیت اولیه، تعدادی از کروموزوم‌ها به‌صورت کاملا تصادفی تولید می‌شوند. در ادامه، مراحل اصلی در طراحی *GA* توصیف می‌شود.

هدف از این بهینه‌سازی، پیدا کردن ترکیبی از ۱۵ ویژگی (در سه مرکز خوشه) از میان ۱۵×۹۱۲۰۰۱ ویژگی نامزد به گونه‌ای که بیشترین فاصله بیرون خوشه‌ای در بین سه خوشه بوجود آید. بنابراین تابع بهینگی به‌عنوان مجموع فاصله‌های بین سه مرکز خوشه در نظر گرفته شده است. هم‌چنین از فاصله اقلیدسی به عنوان معیار عدم شباهت استفاده شده است. در رابطه (۱) تابع هدف مسئله ارائه شده که هدف بیشینه کردن آن است.

کروموزوم	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵

شکل ۶: کدگذاری جواب مسئله در قالب یک کروموزوم

$$\text{Objective Function} = \sqrt{\sum_{i=1}^{5}(p_i - q_i)^2} + \sqrt{\sum_{i=1}^{5}(p_i - h_i)^2} + \sqrt{\sum_{i=1}^{5}(q_i - h_i)^2} \qquad \text{رابطه(۱)}$$

در رابطه (۱)، i شماره ویژگی یک مرکز خوشه، p_i هم i-امین ویژگی از مرکز خوشه اول، q_i نیز i-امین ویژگی از مرکز خوشه دوم و h_i نیز i-امین ویژگی از مرکز خوشه سوم می‌باشند. برای انتخاب والدین هر نسل، از روش چرخ گردان استفاده شده است. طبق رابطه (۲)، کروموزومی که تابع بهینگی بهتری دارد، شانس بیشتری برای انتخاب دارد [۳۴].

رابطه(۲)

$$p_r = \frac{F(X_r)}{\sum_{k=1}^{n} F(X_k)}$$

در رابطه (۲)، P_r احتمال انتخاب کروموزوم r، $F(X_r)$ تابع بهینگی کروموزوم r و n تعداد کل کروموزوم‌ها می‌باشد. هم‌چنین برای عمل تقاطع از ترکیب تک‌نقطه‌ای استفاده شده است [۳۵، ۴۵ و ۴۶]. پس از انتخاب دو والد، ابتدا بر اساس نرخ ترکیب، احتمال اجرای آن بررسی می‌شود و سپس همانند شکل (۷) فرایند ترکیب صورت می‌گیرد. همان‌طور که اشاره شد، سه مرکز خوشه وجود دارد که هر کدام ۵ ویژگی دارند. لذا ترکیب بین دو والد به این صورت است که ۵ ویژگی اول از والد اول با ۵ ویژگی اول از والد دوم، ۵ ویژگی وسط از والد اول با ۵ ویژگی وسط از والد دوم و ۵ ویژگی آخر از والد اول با ۵ ویژگی آخر از والد دوم با هم ترکیب می‌شوند.

والد۱	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
والد۲	۲۰	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴

فرزند۱	۱	۲	۳	۲۳	۲۴	۶	۷	۸	۲۸	۲۹	۱۱	۱۲	۳۲	۳۳	۳۴
فرزند۲	۲۰	۲۱	۲۲	۴	۵	۲۵	۲۶	۲۷	۹	۱۰	۳۰	۳۱	۱۳	۱۴	۱۵

شکل ۷: نمونه‌ای از ترکیب تک‌نقطه‌ای

برای هر کروموزوم حاصل از یک تقاطع، یک عدد تصادفی بین صفر و یک تولید شده و اگر این عدد کوچک‌تر از نرخ جهش باشد، عمل جهش صورت می‌پذیرد. به این صورت که یک ژن از ۵ ژن اول، یک ژن از ۵ ژن دوم و یک ژن از پنج ژن سوم، به‌صورت تصادفی انتخاب می‌شوند و با شماره سه ویژگی از ویژگی‌هایی که در کروموزوم مربوطه نیستند، تعویض خواهند شد. در شکل (۸) نمونه‌ای از فرایند جهش نشان داده شده است.

کروموزوم	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
جهش‌یافته	۱	۲	۱۰۰	۴	۵	۶	۲۰۰	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۳۰۰

شکل ۸: نمونه‌ای از عملگر جهش

تکمیل مراحل فوق، یک اجرای کامل از *GA* می‌باشد و نسل جدیدی از کروموزوم‌ها ایجاد شده است. تمامی جواب‌ها به ترتیب اهمیت تابع بهینگی مرتب می‌شوند. کروموزوم با بهترین تابع بهینگی به‌عنوان بهترین جواب مسئله ذخیره می‌شود. هم‌چنین تعدادی کروموزوم به‌عنوان جواب‌های نخبه وارد نسل بعدی می‌شوند. شرط پایان حلقه الگوریتم تعداد اجرای مشخص می‌باشد. نتایج حاصل از اجرای *GA* با استفاده از پارامترهای مختلف، در جدول (۲) آمده است.

جدول ۲: کالیبره کردن پارامترهای *GA* با روش سعی و خطا

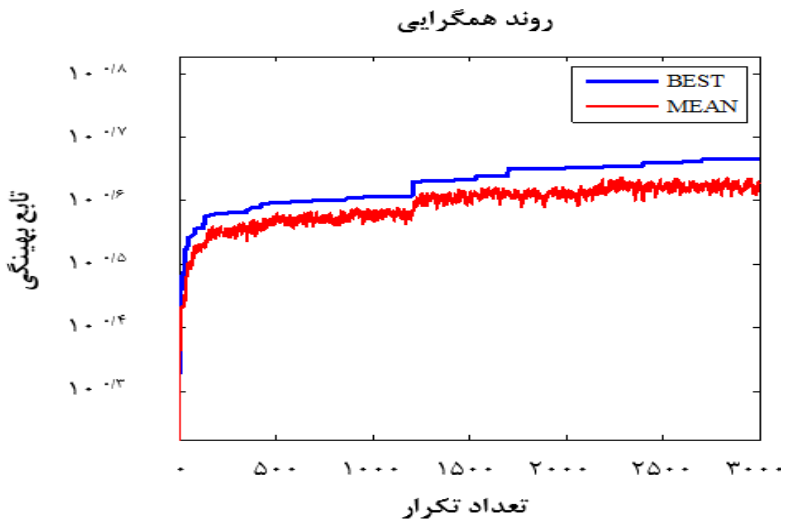
تعداد اجرا	تعداد تکرار	جمعیت اولیه	درصد نخبه‌گرایی	نرخ جهش	تعداد ژن برای ترکیب	زمان اجرا (دقیقه)	تابع بهینگی (بدون واحد)
۱	۲۰۰۰	۲۰۰	۳	۰٫۲۰	[۱ ۳]	۲۸	۴٫۰۰۱۴
۲	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۳]	۲۷	۴٫۲۴۵۶
۳	۲۰۰۰	۲۰۰	۵۰	۰٫۲۰	[۱ ۳]	۲۹	۳٫۸۴۵۶
۴	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۳]	۲۶	۴٫۲۴۵۱
۵	۲۰۰۰	۲۰۰	۱۰	۰٫۶۰	[۱ ۳]	۲۸	۴٫۰۲۴۱
۶	۲۰۰۰	۲۰۰	۱۰	۰٫۰۲	[۱ ۳]	۲۷	۳٫۲۲۶۷
۷	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۵]	۲۸	۳٫۹۴۱۵
۸	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۴]	۲۶	۴٫۰۲۰۱
۹	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۳]	۲۹	۴٫۲۰۸۵
۱۰	۲۰۰۰	۱۰۰	۱۰	۰٫۲۰	[۱ ۳]	۱۴	۳٫۹۷۶۵
۱۱	۲۰۰۰	۲۰۰	۱۰	۰٫۲۰	[۱ ۳]	۲۷	۴٫۲۱۲۴
۱۲	۲۰۰۰	۳۰۰	۱۰	۰٫۲۰	[۱ ۳]	۴۸	۴٫۳۹۵۶
۱۳	۲۰۰۰	۴۰۰	۱۰	۰٫۲۰	[۱ ۳]	۶۳	۴٫۴۴۲۱
۱۴	۱۵۰۰	۳۰۰	۱۰	۰٫۲۰	[۱ ۳]	۳۴	۴٫۱۹۰۱
۱۵	۲۰۰۰	۳۰۰	۱۰	۰٫۲۰	[۱ ۳]	۴۶	۴٫۳۱۲۵
۱۶	۲۵۰۰	۳۰۰	۱۰	۰٫۲۰	[۱ ۳]	۵۷	۴٫۵۶۴۷
۱۷	۳۰۰۰	۳۰۰	۱۰	۰٫۲۰	[۱ ۳]	۶۸	۴٫۶۱۸۳

برای پیدا کردن بهترین ارزش برای یک پارامتر، پارامترهای دیگر را ثابت نگه داشته و الگوریتم با مقادیر مختلف آن پارامتر اجرا شده است. تغییرات در مقادیر سه پارامتر درصد نخبه‌گرایی، نرخ جهش و تعداد ژن‌های جایگزین شده برای ترکیب تأثیر مهمی روی زمان اجرای الگوریتم نمی‌گذارند. بنابراین مقدار تابع بهینگی به‌عنوان معیارهای اصلی برای اندازه‌گیری و کالیبراسیون پارامترهای الگوریتم در نظر گرفته شده است. مسئله مهم در ارزیابی این گونه الگوریتم‌ها این است که تا چه حدی موفق به یافتن جواب‌های بهتر و بهینه می‌گردند. لذا معیار مهم و مورد استفاده در این تحقیق برای تنظیم پارامترهای کالیبراسیون و ارزیابی الگوریتم، همان بهترین مقادیر تابع بهینگی یافت شده توسط الگوریتم لحاظ شده است. در این روش، برای هر یک از پارامترهای کالیبراسیون، تعداد مقادیر بسیار متنوعی انتخاب و تست شده‌اند. ولی به اختصار و برای نمایش آن‌ها در جدول تنها تعداد محدودی از حالت‌ها لحاظ شده‌اند و نتایج آن‌ها ارائه داده شده است. نرخ جهش ۰٫۲۰ بهترین نتایج را تولید کرده است. سطح پایین نخبه‌گرایی ازجمله سه درصد، سرعت همگرایی الگوریتم را کاهش می‌دهد. در مقابل، سطوح بالاتر آن ازجمله ۵۰ درصد، باعث می‌شود الگوریتم در بهینه محلی قرار بگیرد. بهترین حالت نخبه‌گرایی ۱۰ درصد بوده است. هم‌چنین تعداد ژن‌های جایگزین شده برای عمل ترکیب در حالت یک تا سه ژن بهترین نتایج را ارائه داده است.

با جمعیت اولیه برابر با ۱۰۰ کروموزوم، الگوریتم جواب‌های مناسبی پیدا نکرده است. با افزایش تعداد جمعیت به ۲۰۰ کروموزوم، الگوریتم بهبود قابل ملاحظه‌ای داشته است، ولی هم‌چنان جواب بهینه

عمومی کشف نشده است. در نهایت با جمعیت ۳۰۰ کروموزوم الگوریتم توانسته است جواب بهینه را تعیین کند. تغییر تعداد جمعیت از ۳۰۰ به ۴۰۰ کروموزوم چندان تأثیر نمی‌گذارد. بنابراین جمعیت اولیه برابر با ۳۰۰ کروموزوم که منجر به یک پیاده‌سازی سریع‌تر می‌باشد، انتخاب شده است. در تکرار ۱۵۰۰، الگوریتم از جواب بهینه فاصله دارد. در تکرار ۲۰۰۰، الگوریتم به جواب بهینه نزدیک شده است و در تکرارهای ۲۵۰۰ و ۳۰۰۰ الگوریتم نتایج بهتری را ارائه داده است. با توجه به اینکه فضای جستجوی الگوریتم بسیار بالا می‌باشد، لذا هرچقدر پارامترهای جمعیت اولیه و تعداد تکرار بالاتر تنظیم شوند، طبیعی است که الگوریتم نتایج بهتری را ارائه می‌دهد. از طرفی زمان اجرا نیز بالاتر خواهد رفت. در واقع این موضوع ذات الگوریتم‌های فراابتکاری را نشان می‌دهد. به این معنی که این الگوریتم‌ها یا به جواب بهینه مسئله دست خواهند یافت و یا حداقل به پاسخ بهینه نزدیک می‌شوند.

شکل (۹) روند همگرایی *GA* در یک اجرای خاص را نشان می‌دهد. همان‌طور ک مشاهده می‌شود، نمودار آبی رنگ بهترین مقدار تابع بهینگی کروموزوم‌ها در هر نسل را نشان می‌دهد که در نهایت در تکرار ۳۰۰۰ بهترین جواب بدست آمده توسط الگوریتم بدست آمده است. نمودار قرمز رنگ نیز میانگین تابع بهینگی تمامی کروموزوم‌های یک نسل (یک تکرار) را نشان می‌دهد. از آن‌جایی که کروموزوم‌های موجود در یک نسل مقادیر تابع بهینگی متفاوتی دارند، لذا روند میانگین تابع بهینگی به صورت نوسانی تغییر می‌کند. بهترین مقدار تابع برازندگی (در اجراهای مختلف و با تکرار پذیری بالا) برابر با ۴٫۶۱۸۳ بوده است (بدون واحد). ویژگی‌ها با شماره‌های {۲۸۵۵۶۴، ۲۹۲۰۶۸، ۲۹۵۹۷۴، ۲۹۸۵۴۹، ۳۰۹۰۵۸، ۳۰۹۰۸۵، ۳۱۵۵۹۶، ۳۱۶۸۲۶، ۳۱۸۱۱۰، ۴۵۷۴۵۷، ۴۶۶۴۷۶، ۴۶۷۷۷۵، ۷۳۲۷۳۶، ۷۵۱۰۱۱ و ۷۸۵۵۹۵}، ویژگی‌های می‌باشند که توسط *GA* یافت شده‌اند. این ۱۵ ویژگی که سه مرکز خوشه بهینه را نتیجه می‌دهند، ورودی الگوریتم *K-Means* می‌باشند.



شکل ۹: روند همگرایی *GA*

۳-۲- پیاده‌سازی الگوریتم *K-Means* مبتنی بر *GA*

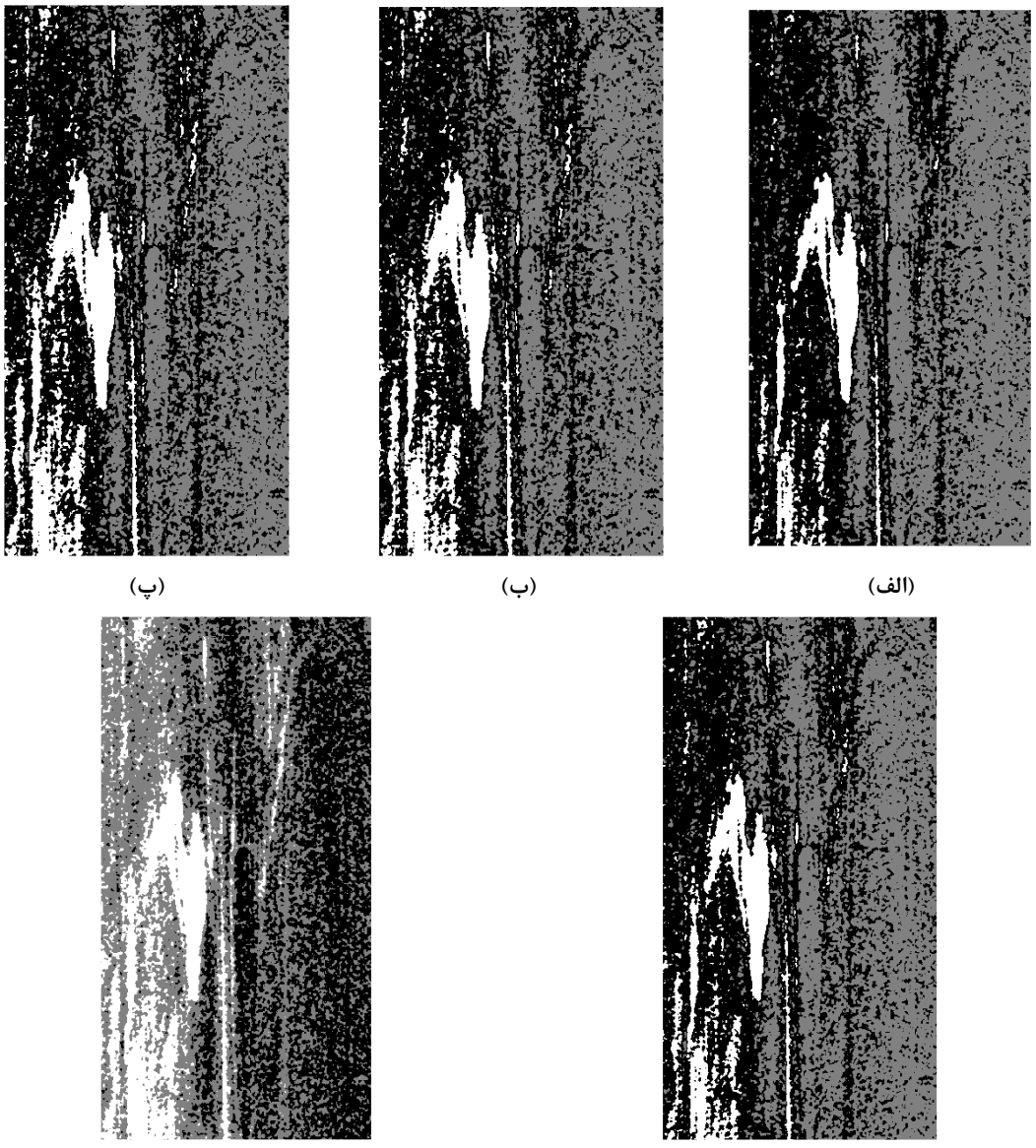
در این تحقیق جهت ارزیابی روش خوشه‌بندی *K-Means* بهبودیافته با *GA* از الگوریتم‌های بهینه‌سازی ازدحام ذرات (*PSO*)، بهینه‌سازی مبتنی بر جغرافیای زیستی (*BBO*)، الگوریتم کلونی زنبور مصنوعی (*ABC*) و روش خوشه‌بندی *K-Means* استاندارد استفاده شده است. پارامترهای کالیبراسیون الگوریتم‌های پیشنهادی در جدول (۳) آمده است.

جدول ۳: تنظیم پارامترهای الگوریتم‌های بهینه‌ساز با روش سعی و خطا (تعداد تکرار=۳۰۰۰، جمعیت اولیه=۳۰۰)

الگوریتم	پارامتر	مقدار
PSO	نرخ حرکت اینرسی (α)	۰/۱۲
	نرخ حرکت به سمت بهترین تجربه شخصی (*Φ1*)	۰/۷۱
	نرخ حرکت به سمت بهترین تجربه همسایه‌ها (*Φ2*)	۰/۸۶
BBO	بازه احتمال مهاجرت به هر ژن	[۰ ۱]
	ماکزیمم نرخ مهاجرت به داخل و مهاجرت به بیرون	۱
	درصد نخبه‌گرایی	٪۱۲
	نرخ جهش	۰/۱۵
ABC	تعداد زنبورهای پیشاهنگ	۳۰۰
	تعداد زنبورهای تماشاگر	۲۰۰
	تعداد زنبورهای مبلغ	۱۶۰

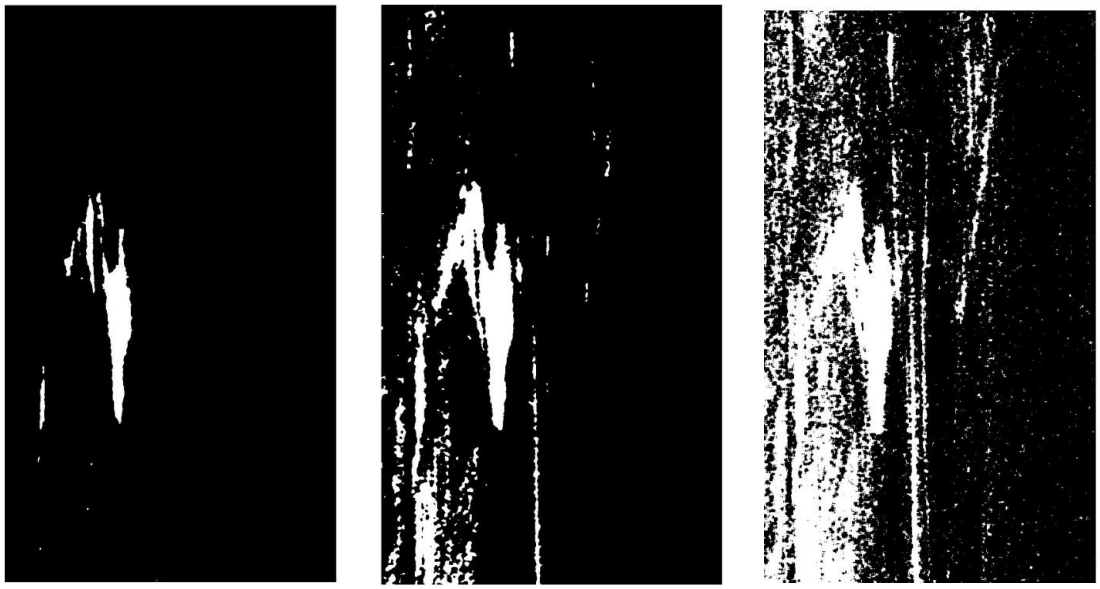
جهت خوشه‌بندی تصاویر به سه کلاس دریا، *look-alike* و لکه‌های نفتی، از الگوریتم *K-Means* استفاده شده است. به این صورت که مقادیر بهینه مراکز خوشه به دست آمده توسط الگوریتم‌های بهینه‌ساز (*GA*، *PSO*، *BBO* و *ABC*)، به عنوان مراکز ورودی در الگوریتم *K-Means* به کار گرفته شده است.

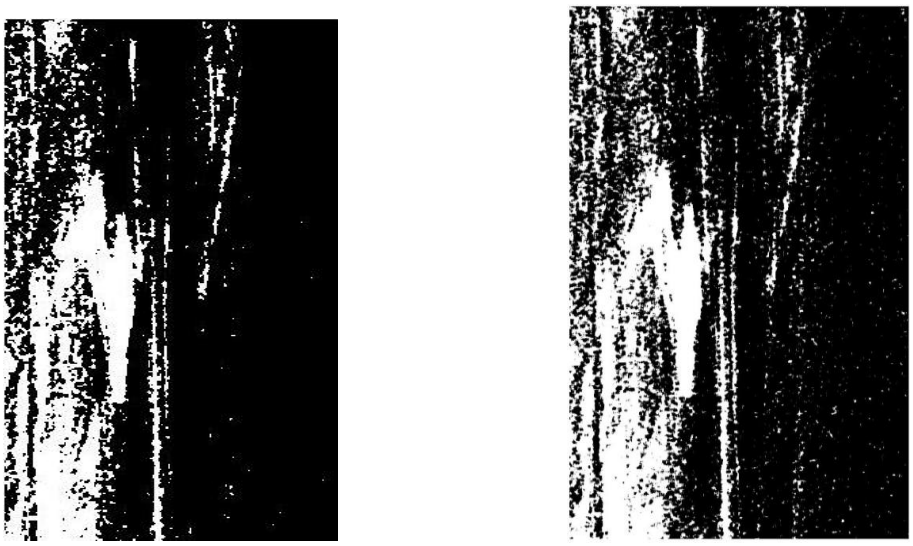
در شکل (۱۰-الف) تصویر خروجی خوشه‌بندی لکه‌های نفتی و غیر نفتی حاصل از روش *K-Means* مبتنی بر *GA* آمده است. همان‌طور که در شکل (۱۰-الف) مشاهده می‌شود رنگ‌های سفید در تصویر بیانگر لکه‌های نفتی در منطقه مطالعاتی می‌باشند. رنگ‌های مشکی بیان‌گر پدیده *look-alike* و رنگ‌های طوسی هم بیان‌گر کلاس دریا می‌باشند. هم‌چنین شکل (۱۰-ب)، تصویر خروجی خوشه‌بندی لکه‌های نفتی و غیر نفتی حاصل از روش *K-Means* معمولی را نشان می‌دهد. با توجه به داده‌های واقعیت زمینی، روش *K-Means* مبتنی بر *GA* نتایج مناسب‌تری را ارائه داده است. همچنین روش *K-Means* مبتنی بر *ABC* نتایج بهتری را نسبت به الگوریتم‌های *PSO-K-Means* و *BBO-K-Means* بدست آورده است. قابل ذکر می‌باشد که در تصاویر خروجی شکل(۱۰)، ویژگی‌های هر پنج تصویر به صورت هم‌زمان در نظر گرفته شده است.



شکل ۱۰: تصویر خروجی خوشه‌بندی با در نظر گرفتن پنج ویژگی؛ الف) روش *K-Means* مبتنی بر *GA*، ب) روش *K-Means* مبتنی بر *PSO*، پ) روش *K-Means* مبتنی بر *BBO*، ت) از روش *K-Means* مبتنی بر *ABC*، ث) *K-Means* استاندارد

هم‌چنین تصویر خروجی خوشه‌بندی لکه‌های نفتی حاصل از هر یک از ویژگی‌های موردنظر (هر تصویر به صورت جداگانه) در شکل (۱۱) نشان داده شده است. تمامی ویژگی‌هایی که به عنوان ویژگی کارآمد در بخش پیش‌پردازش داده‌ها تعریف شده‌اند، به عنوان ورودی به الگوریتم معرفی شده‌اند. از آن‌جایی‌که هر یک از ویژگی‌های استخراج شده دارای یک محدوده عددی خاصی می‌باشند، لذا تمامی مقادیر ویژگی‌ها در بازه صفر و یک نرمالایز شده‌اند.





شکل ۱۱: تصاویر خروجی شناسایی لکه‌های نفتی برای ویژگی‌های مختلف به صورت جداگانه

۳-۳- ارزیابی خوشه‌بندی

همان‌طور که در بخش ۲ اشاره شد، به منظور ارزیابی عملکرد هر یک از ویژگی‌های معرفی شده، از داده واقعیت زمینی که از رقومی‌سازی قسمت مشخصی از لکه نفتی به دست آمده، بهره گرفته شده است. در این قسمت نتایج حاصل از الگوریتم‌های بهبودیافته و الگوریتم *K-Means* استاندارد و هم‌چنین تاثیر هر یک از ویژگی‌ها بر دقت خوشه‌بندی مورد مقایسه قرار گرفته‌اند. برای ارزیابی دقیق‌تر از معیارهای دقت کلی[1]، ضریب کاپا[2]، خطای نوع اول[3] و خطای نوع دوم[4] بهره گرفته شده است که این معیارها به صورت زیر تعیین می‌شوند:

[1] *Overall Accuracy (OA)*
[2] *Kappa Coefficient (KC)*
[3] *First Type Error (FTE)*
[4] *Second Type Error (STE)*

- دقت کلی: نسبت تعداد پیکسل‌های درست تشخیص داده شده، به تعداد کل پیکسل‌های کلاس مورد نظر
- خطای نوع اول: نسبت پیکسل‌های لکه نفتی که به اشتباه در کلاس دریا قرار گرفته‌اند، به پیکسل‌های کلاس دریا
- خطای نوع دوم: نسبت پیکسل‌های دریا که به اشتباه در کلاس لکه نفتی قرار گرفته‌اند، به پیکسل‌های لکه نفتی

معیارهای دقت و صحت حاصل از روش‌های *K-Means* مبتنی بر *GA*، *ABC-K-Means*، *PSO-K-Means*، *BBO-K-Means* و *K-Means* استاندارد (مطابق با شکل(۱۰)) در جدول (۴) آمده است. همان‌طور که مشاهده می‌شود الگوریتم پیشنهادی توانسته است دقت کلی ۹۲٫۲۸ را حاصل کند که در قیاس با *K-Means* استاندارد و سایر الگوریتم‌های پیشنهادی از دقت کلی بالاتری برخوردار است. با توجه به خطای نوع دوم حاصل شده می‌توان فهمید، الگوریتم پیشنهادی *GA_K-Means* دارای قطعیت و صحت بالاتری است. یعنی پیکسل‌هایی که متعلق به کلاس دریا بوده‌اند کمتر به عنوان کلاس لکه نفتی نسبت داده شده‌اند.

هم‌چنین طبق نتایج حاصل از شکل (۱۱)، هر یک از ویژگی‌ها عملکرد متفاوتی را در شناسایی لکه نفتی از خود نشان داده‌اند که این اختلافات به صورت عددی در جدول (۵) و به صورت بصری در شکل (۱۲) نشان داده شده‌اند.

جدول ۴: نتایج عددی حاصل شده از خوشه‌بندی *K-Means* مبتنی بر *GA* و سایر الگوریتم‌های پیشنهادی

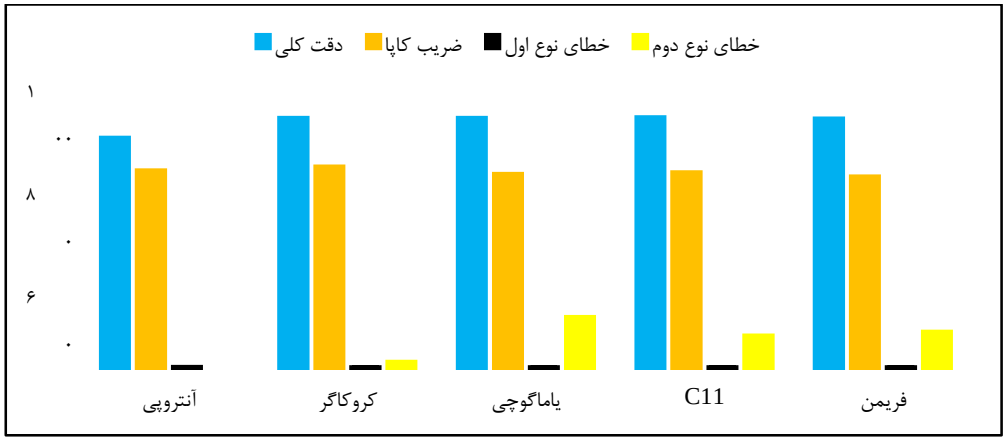
الگوریتم	دقت کلی	خطای نوع اول	خطای نوع دوم	ضریب کاپا
GA_K-Means	۹۲٫۲۸	<۰٫۰۲	۰٫۱۱	۷۴٫۱۳
PSO_K-Means	۹۰٫۱۲	<۰٫۰۱	۲٫۹۱	۷۱٫۳۲
BBO_K-Means	۹۰٫۲۶	<۰٫۰۲	۳٫۰۹	۷۱٫۸۴
ABC_K-Means	۹۰٫۸۴	<۰٫۰۲	۲٫۳۶	۷۲٫۱۶
K-Means	۸۹٫۲۴	<۰٫۰۱	۶٫۴۲	۷۰٫۶۷

جدول ۵: نتایج عددی حاصل شده از خوشه‌بندی لکه‌های نفتی برای هر یک از ویژگی‌ها

ویژگی	دقت کلی	خطای نوع اول	خطای نوع دوم	ضریب کاپا
آنتروپی	۸۳٫۲۴	۰٫۰۲	۰	۷۱٫۴۱
کروکاگر	۹۰٫۳۶	<۰٫۰۱	۲٫۰۴	۷۲٫۷۷
یاماگوچی	۹۰٫۴۴	<۰٫۰۱	۱۸٫۳۰	۷۰٫۱۱
C11	۹۰٫۵۷	<۰٫۰۱	۱۱٫۵۴	۷۰٫۷۱
فریمن	۸۹٫۲۰	<۰٫۰۱	۱۲٫۹۰	۶۹٫۱۷

همانطور که در شکل (۱۲) و جدول (۴) ملاحظه می‌شود، ویژگی آنتروپی توانسته است دقت کلی ۸۳٫۲۴ را حاصل کند که در قیاس با سایر ویژگی‌ها از دقت کلی کمتری برخوردار است. اما با توجه به خطای نوع اول و دوم حاصل شده از این ویژگی می‌توان فهمید، اگرچه دقت کلی کمتر از سایر دقت‌های کلی ارائه شده می‌باشد، اما دارای قطعیت و صحت بالاتری است (نویز کمتری دارد). یعنی پیکسل‌هایی که متعلق به کلاس دریا بوده‌اند کمتر به عنوان کلاس لکه نفتی نسبت داده شده‌اند. هم‌چنین ویژگی مولفه *Ks* تجزیه کننده کروگاگر، به نسبت سه ویژگی دیگر، حالت بهینه‌تری را از خود در مقادیر عددی نشان داده است. به طوریکه

دقت کلی قابل قبول ۹۰ درصدی را حاصل نموده و دچار خطای نوع دوم کمتری نسبت به سایر ویژگی‌ها شده است. همان‌طور که از شکل (۱۱) برمی‌آید، نسبت به ویژگی آنتروپی، پیکسل‌هایی بیشتری از دریا را به عنوان کلاس لکه نفتی شناسایی کرده است، اما نسبت به سه ویژگی دیگر پیکسل‌های کمتری از کلاس دریا را به عنوان کلاس لکه نفتی شناسایی کرده است. هم‌چنین هر یک از ویژگی‌های یاماگوچی، فریمن و مولفه *C11* از ماتریس کواریانس علارغم اینکه دقت کلی ۹۰ درصدی را حاصل کرده‌اند، اما به ترتیب با خطای نوع دوم برابر با ۱۸، ۱۱ و ۱۲ درصدی، صحت کمتری را در شناسایی لکه نفتی از خود نشان داده‌اند. برای تصویر یاماگوچی مقدار خطای نوع دوم بسیار بالاتر از سایر ویژگی‌ها می‌باشد که این نشان دهنده تحت تاثیر بودن الگوریتم پیشنهادی به مقادیر نویز موجود در تصویر و فضای ویژگی می‌باشد.



شکل ۱۲: مقایسه بصری نتایج آماری حاصل شده از خوشه‌بندی لکه‌های نفتی برای هر یک از ویژگی‌ها

۴- نتیجه‌گیری و پیشنهادات

در این تحقیق از روش *K-Means* مبتنی بر الگوریتم *GA* جهت شناسایی لکه‌های نفتی در سطح دریا استفاده شده است. به‌طور کلی نتایج پیاده‌سازی الگوریتم *K-Means* بهبودیافته با *GA* برای شناسایی لکه‌های نفتی، قابلیت اجرای بالایی را نشان داده است. به منظور ارزیابی عملکرد تصاویر حاصل از الگوریتم‌ها، از داده واقعیت زمینی دیجیتایز شده استفاده شده است. مطابق با نتایج، در برخی از ویژگی‌ها از جمله ویژگی‌های *ODD* تجزیه‌کننده یاماگوچی، *ODD* تجزیه کننده فریمن و مولفه *C11* از ماتریس کواریانس، دقت کلی مناسب بوده است اما خطای نوع دوم چندان مناسب نبوده است. در واقع اگرچه این ویژگی‌ها دقت کلی قابل قبولی را از خود نشان داده‌اند اما بایستی در نظر گرفته شود که پیکسل‌های دریای زیادی را به اشتباه در کلاس لکه نفتی قرار می‌دهند که نمی‌تواند ایده‌ال یک کاربر باشد. بنابراین صحت خوشه‌بندی از اهمیت بسیار بالایی برخوردار می‌باشد.

هم‌چنین در این تحقیق الگوریتم *K-Means* با استفاده از الگوریتم فرااکتشافی ژنتیک بهبود یافته است. در حالت کلی، توسعه یک الگوریتم فرااکتشافی، باید قبل از اجرا و پیاده‌سازی و بر مبنای یک منطق ریاضی مشخص باشد. این منطق می‌تواند از طریق شناخت کامل الگوریتم، نقاط ضعف و قوت آن به دست آید. در این تحقیق با توجه به شعار بیشترین فاصله بیرون خوشه‌ای و بیشترین شباهت درون خوشه‌ای، الگوریتم *K-Means* بهبود داده شده است. هم‌چنین نتایج پیاده‌سازی الگوریتم پیشنهادی نیز دقت کلی و صحت خوبی را به نمایش گذاشته است. بنابراین می‌توان گفت توسعه الگوریتم *K-Means* قابل اعتماد می‌باشد.

در این مطالعه ماکزیمم کردن فاصله بیرون خوشه‌ای به عنوان تابع هدف در *GA* در نظر گرفته شده است. در نظر گرفتن اهداف دیگر از جمله بیشترین شباهت درون خوشه‌ای و استفاده از روش‌های بهینه‌سازی چندهدفه مطمئناً می‌توانند نتایج مناسب‌تری را برای این مسئله به همراه داشته باشند. هم‌چنین در نظر گرفتن ویژگی‌های دیگر تصاویر *SAR* و تاثیر هر یک از آن‌ها بر دقت خوشه‌بندی لکه‌های نفتی، می‌تواند جنبه‌های جدیدی را بر این تحقیق اضافه کند.

مراجع

[1] Z. Jiao, G. Jia and Y. Cai, "A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles," Computers & Industrial Engineering, vol. 135, pp. 1300-1311, 2018.

[2] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: an accurate oil spill detection method based on level set method from synthetic aperture radar imagery," European Journal of Remote Sensing, pp. 1-18, 2022.

[3] J. Xu, X. Pan, B. Jia, X. Wu and B. Li, "Oil spill detection using LBP feature and K-means clustering in shipborne radar image," Journal of Marine Science and Engineering, vol. 9, no. 1, pp. 65, 2021.

[4] W. Alpers, B. Holt and K. Zeng, "Oil spill detection by imaging radars: Challenges and pitfalls," Remote Sensing of Environment, vol. 201, pp. 133-147, 2017.

[5] M. O. Jeffries, K. Morris, W. F. Weeks and H. Wakabayashi, "Structural and stratigraphie features and ERS 1 synthetic aperture radar backscatter characteristics of ice growing on shallow lakes in NW Alaska, winter 1991–1992," Journal of Geophysical Research: Oceans, vol. 99, no. 11, pp. 22459-22471, 1994.

[6] S. Naz, M. F. Iqbal, I. Mahmood and M. Allam, "Marine oil spill detection using Synthetic Aperture Radar over Indian Ocean," Marine Pollution Bulletin, vol. 162, pp. 111921, 2021.

[7] D. Mera, M. Fernández-Delgado, J. M. Cotos, J. R. R. Viqueira and S. Barro, "Comparison of a massive and diverse collection of ensembles and other classifiers for oil spill detection in SAR satellite images," Neural Computing and Applications, vol. 28, no. 1, pp. 1101-1117, 2017.

[8] Y. Guo and H. Z. Zhang, "Oil spill detection using synthetic aperture radar images and feature selection in shape space," International Journal of Applied Earth Observation and Geoinformation, vol. 30, pp. 146-157, 2014.

[9] B. Fiscella, A. Giancaspro, F. Nirchio, P. Pavese and P. Trivero, "Oil spill detection using marine SAR images," International Journal of Remote Sensing, vol. 21, no. 18, pp. 3561-3566, 2000.

[10] F. Nirchio, M., Sorgente, A. Giancaspro, W. Biamino, E. Parisato, R. Ravera and P. Trivero, "Automatic detection of oil spills from SAR images," International Journal of Remote Sensing, vol. 26, no. 6, pp. 1157-1174, 2005.

[11] A. S. Solberg, G. Storvik, R. Solberg and E. Volden, "Automatic detection of oil spills in ERS SAR images," IEEE Transactions on geoscience and remote sensing, vol. 37, no. 4, pp. 1916-1924, 1999.

[12] K. Topouzelis, V. Karathanassi, P. Pavlakis and D. Rokos, "Dark formation detection using neural networks," International Journal of Remote Sensing, vol. 29, no. 16, pp. 4705-4720, 2008.

[13] F. Frate, A. Petrocchi, J. Lichtenegger and G. Calabresi, "Neural networks for oil spill detection using ERS-SAR data," IEEE Transactions on geoscience and remote sensing, vol. 38, no. 5, pp. 2282-2287, 2000.

[14] T. F. Kanaa, E. Tonye, G. Mercier, V. D.

P. Onana and J. P. Rudant, "Multiscale Segmentation of Oil Slick in SAR Images based on Morphological Pyramid," In ENVISAT and ERS Symposium, Salsburg, Australie, pp. 6-10, 2004.

[15] A. K. Liu, C. Y. Peng and S. S. Chang, "Wavelet analysis of satellite images for coastal watch," IEEE Journal of Oceanic Engineering, vol. 22, no. 1, pp. 9-17, 1997.

[16] S. Y. Wu and A. K. Liu, "Towards an automated ocean feature detection, extraction and classification scheme for SAR imagery," International Journal of Remote Sensing, vol. 24, no. 5, pp. 935-951, 2003.

[17] S. Derrode and G. Mercier, "Unsupervised multiscale oil slick segmentation from SAR images using a vector HMC model," Pattern Recognition, vol. 40, no. 3, pp. 1135-1147, 2007.

[18] R. T. Araújo, F. N. de Medeiros, R. C. Costa, R. C. Marques, R. B. Moreira and J. L. Silva, "Locating oil spill in SAR images using wavelets and region growing," In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Heidelberg, pp. 1184-1193, 2004.

[19] G. Benelli and A. Garzelli, "Oil-spills detection in SAR images by fractal dimension estimation," In International Geoscience and Remote Sensing Symposium, vol. 1, pp. 218-220, 1999.

[20] M. Marghany, M. Hashim and A. P. Cracknell, "Fractal dimension algorithm for detecting oil spills using RADARSAT-1 SAR," In International Conference on Computational Science and Its Applications, Springer, Berlin, Heidelberg, pp. 1054-1062, 2007.

[21] M. Marghany, A. P. Cracknell and M. Hashim, "Modification of fractal algorithm for oil spill detection from RADARSAT-1 SAR data," International Journal of Applied Earth Observation and Geoinformation, vol. 11, no. 2, pp. 96-102, 2009.

[22] D. Latini, F. Del Frate and C. E. Jones, "Multi-frequency and polarimetric quantitative analysis of the Gulf of Mexico oil spill event comparing different SAR systems," Remote sensing of environment, vol. 183, pp. 26-42, 2016.

[23] Y. Li and J. Li, "Oil spill detection from SAR intensity imagery using a marked point process," Remote Sensing of Environment, vol. 114, no. 7, pp. 1590-1601, 2010.

[24] S. Skrunes, C. Brekke and T. Eltoft, "Oil spill characterization with multi-polarization C-and X-band SAR," In International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5117-5120, 2012.

[25] K. A. S. Al Abri, N. Poojary and J. Menezes, "A Novel Segmentation Technique for Clustering Oil Spill Data from Hyperspectral Images," Proc. of Int. Conf. on Advances in Communication and Information Technology, 2012.

[26] J. Xiao, Y. Yan, J. Zhang and Y. Tang, "A quantum-inspired genetic algorithm for k-means clustering," Expert Systems with Applications, vol. 37, no. 7, pp. 4966-4973, 2010.

[27] F. G. Tari and Z. Hashemi, "Prioritized K-mean clustering hybrid GA for discounted fixed charge transportation problems," Computers & Industrial Engineering, vol. 126, pp. 63-74, 2018.

[28] M. Kaveh, M. S. Mesgari and A. Khosravi, "Solving the local positioning problem using a four-layer artificial neural network," Engineering Journal of Geospatial Information Technology, vol. 7, no. 4, pp. 21-40, 2020.

[29] M. Z. Islam, V. Estivill-Castro, M. A. Rahman and T. Bossomaier, "Combining K-Means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering," Expert Systems with Applications, vol. 91, pp. 402-417, 2018.

[30] M. Kaveh, M. Kaveh, M. S. Mesgari and R. Sadeghi Paland, "Multiple criteria decision-making for hospital location-allocation based on improved genetic algorithm," Applied Geomatics, vol. 12, no. 3, pp. 291-306, 2020.

[31] K. J. Kim and H. Ahn, "A recommender system using GA K-means clustering in an online shopping market," Expert systems with applications, vol. 34, no. 2, pp. 1200-1209, 2008.

[32] A. A. Matkan, M. Hajeb and Z. Azarakhsh, "Oil spill detection from SAR image using SVM based classification," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, SMPR, vol. 1, pp. W3, 2013.

[33] M. Kaveh and M. S. Mesgari, "Improved biogeography-based optimization using migration process adjustment: an approach for location-allocation of ambulances," Computers & Industrial Engineering, vol. 135, p. 800-813, 2019.

[34] O. Rostami and M. Kaveh, "Optimal feature selection for SAR image classification using biogeography-based optimization (BBO), artificial bee colony (ABC) and support vector machine (SVM): a combined approach of optimization and machine learning," Computational Geosciences, vol. 25, no. 3, pp. 911-930, 2021.

[35] J. Wang, M. Khishe, M Kaveh and H. Mohammadi, "Binary chimp optimization algorithm (BChOA): a new binary meta-heuristic for solving optimization problems," Cognitive Computation, vol. 13, no. 5, pp.1297-1316, 2021.

[36] S. Sun, R. Liu, C. Yang, H. Zhou, J. Zhao and J. Ma, "Comparative study on the speckle filters for the very high-resolution polarimetric synthetic aperture radar imagery," Journal of Applied Remote Sensing, vol. 10, no. 4, pp.045014, 2016.

[37] S. Foucher and C. López-Martínez, "Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques," IEEE transactions on image processing, vol. 23, no. 4, pp. 1751-1764, 2014.

[38] J. S. Lee, T. L. Ainsworth and Y. Wang, "On polarimetric SAR speckle filtering," In 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 111-114, 2012.

[39] F. Sadeghi, A. Larijani, O. Rostami, D. Martín and P. Hajirahimi, "A Novel Multi-Objective Binary Chimp Optimization Algorithm for Optimal Feature Selection: Application of Deep-Learning-Based Approaches for SAR Image Classification," Sensors, vol. 23, no. 3, pp. 1180, 2023.

[40] M. Kaveh, M. S. Mesgari and B. Saeidian, "Orchard Algorithm (OA): A new meta-heuristic algorithm for solving discrete and continuous optimization problems," Mathematics and Computers in Simulation, vol. 208, pp. 95-135, 2023.

[41] M. Kaveh, M. S. Mesgari, D. Martín and M. Kaveh, "TDMBBO: a novel three-dimensional migration model of biogeography-based optimization (case study: facility planning and benchmark problems)," The Journal of Supercomputing, vol. 2023, pp. 1-56, 2023.

[42] M. Kaveh and M. S. Mesgari, "Application of meta-heuristic algorithms for training neural networks and deep learning architectures: a comprehensive review," Neural Processing Letters, vol. 2022, pp. 1-104, 2022.

[43] S. Baniasadi, O. Rostami, D. Martín and M. Kaveh, "A Novel Deep Supervised Learning-Based Approach for Intrusion Detection in IoT Systems," Sensors, vol. 22, no. 12, pp. 4459, 2022.

[44] F. Sadeghi, O. Rostami, M. K. Yi and S. Hwang, "A deep learning approach for detecting Covid-19 using the chest X-ray images," CMC-Computers Materials & Continua, vol. 74, no. 1, pp. 751-768, 2023.

[45] M. Kaveh and M. S. Mesgari, "Hospital site selection using hybrid PSO algorithm-

Case study: District 2 of Tehran," Scientific-Research Quarterly of Geographical Data (SEPEHR), vol. 28, no. 111, pp. 7-22, 2019.

[46] N. Kianfar, M. S. Mesgari, A. Mollalo and M. Kaveh, "Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms," Spatial and Spatio-temporal Epidemiology, vol. 40, pp. 100471, 2022.

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.12, No.2, Summer 2024

Research Paper

Improvement of K-Means clustering algorithm using genetic algorithm for spatial analysis of the oil spill detection in polarimetric SAR images

Mehrdad Kaveh [1], Yaser Ebrahimian Ghajari [2]*

1- Ph.D. Student of Geographic Information System in Department of Geodesy and Geomatics, K. N. Toosi University of Technology
2- Assistant Professor in Department of Civil Engineering, Babol Noshirvani University of Technology

Abstract

The existence of the oil spills at the bottom of the ocean and sea is one of main concerns for researchers in marine ecosystem terrains. In this study, K-Means clustering based on genetic algorithm (GA) has been used in order to detect the oil spills at the sea bottom. The main objective of the developed K-Means with GA is to cause an intelligent search which not only randomly determines the initial cluster centers but also tries to find out the optimal ones for the clusters. To achieve this goal, firstly the required preprocessing steps for SAR images such as radiometric correction, speckle reduction and POLSAR feature extraction were applied. Then the optimal cluster centers were determined by GA in order to have maximum extra_cluster distance. Finally, to find out the final optimal centers, K-Means algorithm was used in order to have maximum intra_cluster similarity. The ground truth digitized from Pauli RGB image of POLSAR image was utilized to evaluate the performance of the clustering methods. Furthermore, in order to evaluate the improved K-Means algorithm by GA, particle swarm optimization (PSO), biogeography-based optimization (BBO), artificial bee colony (ABC), and the standard K-Means clustering methods were used. The results gained by the improved GA-K-Means have more validity and precision than the other algorithms. The feature Entropy could obtain the overall accuracy of 83.24% that has a lower accuracy compared with the other features. But it shows a higher validity. The features ODD of Yamagouchi, ODD of Freeman decompositions and C11 of covariance matrix have reached the overall accuracy of 90% but have the noticeable values of the second type error by 18%, 11% and 12% which demonstrate the lowest validity in comparison with the other features.

Key words: *Oil spill, SAR images, feature selection, K-Means, genetic algorithm.*

Correspondence Address *: Department of Civil Engineering, Babol Noshirvani University of Technology, Babol, Iran.*
Tel *: +98 9113544095*
Email*: y.ebrahimian@nit.ac.ir*