

# روشی نوین برای خوشه‌بندی داده‌ها با استفاده از الگوریتم بهینه‌سازی چهارگرگ خاکستری

لاله عجمی بختیاروند و زهرا بهشتی

استخراج دانش از مجموعه‌های داده‌ای و تبدیل آن به یک ساختار قابل فهم برای استفاده‌هایی در آینده می‌باشد [۲]. شکاف موجود بین داده‌ها و اطلاعات، سبب ایجاد نیاز به ابزارهای داده‌کاوی شده است تا داده‌های بی‌ارزش را به دانشی ارزشمند تبدیل کند [۳]. برای این کار الگوریتم‌های متعددی وجود دارد که هر یک برای هدف خاصی کاربرد دارند. خوشه‌بندی از مهم‌ترین الگوریتم‌های داده‌کاوی است و کاربرد بسیاری دارد. الگوریتم‌های خوشه‌بندی، اطلاعاتی را که ویژگی‌های نزدیک به هم و مشابه دارند، در دسته‌هایی که به آنها خوشه گفته می‌شود قرار می‌دهند. اکثر الگوریتم‌ها مراکز خوشه را به صورت تصادفی انتخاب می‌کنند و راه حل نهایی وابستگی به انتخاب اولیه این مراکز دارد [۴]. از این رو امکان دارد به جای همگراشدن به بهینه سراسری به بهینه محلی همگرا شود. مسأله خوشه‌بندی یک مسأله ان‌پی-سخت<sup>۲</sup> [۵] محسوب می‌شود و الگوریتم‌های دقیق<sup>۳</sup> قادر به حل نمونه‌هایی از آن با ابعاد کوچک هستند. از این رو می‌توان از الگوریتم‌های فراابتکاری<sup>۴</sup> که از دسته الگوریتم‌های تقریبی هستند و برای حل مسایل پیچیده با ابعاد بالا کاربرد دارند استفاده کرد [۶] تا [۱۱].

الگوریتم‌های فراابتکاری با انتخاب مراکز خوب در میان حجم وسیعی از داده‌ها، خوشه‌بندی را انجام می‌دهند و از نظر پیچیدگی زمانی و مصرف حافظه، شرایط بهتری را نسبت به الگوریتم‌های دقیق فراهم می‌کنند. از قابلیت جستجوی این الگوریتم‌ها برای جستجوی مراکز خوشه‌های مناسب در فضای ویژگی داده‌شده استفاده می‌گردد. در این راستا، تحقیقات زیادی روی بهبود کیفیت خوشه‌بندی با الگوریتم‌های فراابتکاری انجام شده است و هر روز نیز کارهای جدیدی به آن اضافه می‌گردد. الگوریتم‌هایی مانند بهینه‌سازی کلونی مورچه‌ها (ACO) [۱۲]، کلونی زنبور عسل (ABC) [۱۳]، تکامل تفاضلی (DE) [۱۴]، خفاش (BA) [۱۵]، فاخته (CA) [۱۶]، سیاه‌چاله (BH) [۱۷]، ممتیک بهینه‌سازی گرانش ذرات (MPGO) [۱۸]، جستجوی گرانشی (GSA) [۱۹]، گرگ خاکستری (GWO) [۱۱]، گرده‌افشانی گل (FPA) [۲۰]، جستجوی ارگانسیم هم‌زیست (SOS) [۲۱]، ژنتیک (GA) [۲۲]، بهینه‌سازی ازدحام ذرات (PSO) [۲۳]، شیر مورچه (ALO) [۲۴]، گروه میگوها (KHA) [۲۵]، شبیه‌سازی تبرید (SA) [۲۶]، ازدحام سالپ (SSA) [۲۷] و الگوریتم بهینه‌سازی نهنگ (WOA) [۲۸] از جمله الگوریتم‌های فراابتکاری هستند که در خوشه‌بندی استفاده شده‌اند.

دلیل کاربرد زیاد این الگوریتم‌ها در خوشه‌بندی آن است که این

چکیده: امروزه، خوشه‌بندی داده‌ها به دلیل حجم و تنوع داده‌ها بسیار مورد توجه قرار گرفته است. مشکل اصلی روش‌های خوشه‌بندی معمول این است که در دام بهینه محلی گرفتار می‌آیند. الگوریتم‌های فراابتکاری به دلیل داشتن توانایی فرار از بهینه‌های محلی، نتایج موفق‌تری را در خوشه‌بندی داده‌ها نشان داده‌اند. الگوریتم بهینه‌سازی گرگ خاکستری از جمله این دسته الگوریتم‌ها است که قابلیت بهره‌برداری خوبی دارد و در برخی از مسایل راه حل مناسبی ارائه داده است، اما اکتشاف آن ضعیف است و در بعضی از مسایل به بهینه محلی همگرا می‌شود. در این تحقیق برای بهبود خوشه‌بندی داده‌ها، نسخه بهبودیافته‌ای از الگوریتم بهینه‌سازی گرگ خاکستری به نام الگوریتم بهینه‌سازی چهارگرگ خاکستری ارائه شده که با استفاده از بهترین موقعیت دسته چهارم گرگ‌ها به نام گرگ‌های امگای پیشرو در تغییر موقعیت هر گرگ، قابلیت اکتشاف بهبود می‌یابد. با محاسبه امتیاز هر گرگ نسبت به بهترین راه حل، نحوه حرکت آن مشخص می‌شود. نتایج الگوریتم پیشنهادی چهارگرگ خاکستری با الگوریتم‌های بهینه‌سازی گرگ خاکستری، بهینه‌سازی ازدحام ذرات، کلونی زنبور عسل مصنوعی، ارگانسیم‌های هم‌زیست و بهینه‌سازی ازدحام سالپ در مسأله خوشه‌بندی روی چهارده مجموعه دادگان ارزیابی شده است. همچنین عملکرد الگوریتم پیشنهادی با چند نسخه بهبودیافته از الگوریتم گرگ خاکستری مقایسه شده است. نتایج به دست آمده عملکرد قابل توجه الگوریتم پیشنهادی را نسبت به سایر الگوریتم‌های فراابتکاری مورد مقایسه در مسأله خوشه‌بندی نشان می‌دهد. بر اساس میانگین معیار  $F$  روی تمام مجموعه دادگان، روش پیشنهادی  $82.172\%$  و الگوریتم بهینه ذرات  $78.284\%$  را نشان می‌دهد و در مقایسه با نسخه‌های بهبودیافته الگوریتم گرگ، الگوریتم EGWO که در رتبه بعدی است دارای میانگین معیار  $F$  برابر  $80.656\%$  می‌باشد.

کلیدواژه: الگوریتم‌های فراابتکاری، الگوریتم بهینه‌سازی گرگ خاکستری، الگوریتم بهینه‌سازی چهارگرگ، خوشه‌بندی.

## ۱- مقدمه

با توجه به تولید حجم بالای اطلاعات در زمینه‌های مختلف و این حقیقت که ترکیبی از داده‌های مفید و غیر مفید در اختیار افراد قرار می‌گیرد، لزوم استفاده از روش‌های داده‌کاوی<sup>۱</sup> جهت استخراج اطلاعات مفید از حجم انبوهی از داده‌ها به خوبی احساس می‌گردد [۱]. داده‌کاوی

این مقاله در تاریخ ۱۸ آذر ماه ۱۳۹۹ دریافت و در تاریخ ۲ مهر ماه ۱۴۰۰ بازنگری شد.

لاله عجمی بختیاروند، دانشکده مهندسی کامپیوتر، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران، (email: l\_ajami\_b@sco.iaun.ac.ir).

زهرا بهشتی (نویسنده مسئول)، دانشکده مهندسی کامپیوتر، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران، (email: z-beheshti@iaun.ac.ir).

1. Data Mining

2. NP-Hard

3. Exact Algorithm

4. Meta-Heuristic Algorithm

الگوریتم بهینه‌سازی چهارگرگ خاکستری و استفاده از آن در خوشه‌بندی داده‌ها در بخش ۳ بیان می‌گردد. در ادامه این بخش با استفاده از مثالی، عملکرد روش پیشنهادی برای خوشه‌بندی داده‌ها نشان داده می‌شود. نتایج و بحث در بخش ۴ ارائه می‌گردد و در بخش ۵، نتیجه‌گیری و تحقیقات آینده شرح داده خواهند شد.

## ۲- پژوهش‌ها و کارهای موجود

### ۲-۱ پژوهش‌های اخیر در زمینه خوشه‌بندی با استفاده از الگوریتم‌های فراابتکاری

در یکی از پژوهش‌های اخیر [۳۱]، خوشه‌بندی بر اساس الگوریتم بهینه‌سازی فاخته برای شبکه‌های حسگر بی‌سیم ارائه شده است. مصرف یکنواخت انرژی و بهینه‌سازی آن، یک نگرانی عمده برای طراحی پروتکل خوشه‌بندی و مسیریابی برای شبکه‌های حسگر بی‌سیم در مقیاس بزرگ است. اکثر راه حل‌های مبتنی بر محاسبات و الگوریتم‌های الهام گرفته شده از طبیعت برای مشکل مسیریابی مبتنی بر خوشه برای شبکه سنسور بی‌سیم، دارای مشکل مصرف انرژی نامتوازن می‌باشند، به دلیل آن که گره‌های نزدیک به سینک از نظر بار ترافیکی بیش از حد بارگیری می‌شوند. در تحقیق انجام‌شده، یک روش خوشه‌بندی متوازن انرژی مبتنی بر الگوریتم جستجوی فاخته ارائه شده که از یک تابع هدف جدید برای توزیع یکنواخت سرخوشه‌ها استفاده می‌کند.

با استفاده از الگوریتم‌های کلونی زنبور مصنوعی و تکامل دیفرانسیل، یک روش ترکیبی ارائه گردید که از آن برای ارزیابی بهترین مجموعه سرخوشه‌ها در توازن بار استفاده شده است [۳۲]. برای خوشه‌بندی کارآمد و توازن بار، یک تابع هدف جدید بر اساس میانگین انرژی، فاصله درون خوشه‌ای و پارامترهای تأخیر طراحی گردیده که از نظر میانگین مصرف انرژی، مصرف کل انرژی، انرژی باقیمانده و طول عمر شبکه دارای عملکرد بهتری نسبت به سایر روش‌های مورد مقایسه دارد.

الگوریتم‌های امنیتی زیادی برای محافظت از شبکه‌های تعریف‌شده توسط نرم‌افزار<sup>۵</sup> (SDN) پیشنهاد شده است، با این حال اکثر آنها در برابر حملات مختلف کارایی پایینی دارند. برای این منظور، شکیل و همکاران [۳۳]، یک چارچوب که بتواند ترافیک پویا را اداره کند و شبکه را در برابر حملات DDOS<sup>۶</sup> محافظت کند با استفاده از خوشه‌بندی بر اساس الگوریتم بهینه‌سازی نهنگ طراحی کردند که برای تشخیص درخواست‌ها از نوع حمله مورد استفاده قرار گرفت.

در پژوهشی دیگر در زمینه خوشه‌بندی، سه الگوریتم ازدحام ذرات، ژنتیک و زنبور عسل برای مقاردهی اولیه برای الگوریتم خوشه‌بندی فازی مبتنی بر کرنل استفاده شد [۳۴]. الگوریتم‌های پیشنهادی برای حل یک مطالعه موردی برای تقسیم‌بندی مشتریان در یکی از فروشگاه‌های فروش لباس در تایوان استفاده گردید که نتایج الگوریتم‌های پیشنهادی، ساختار خوشه‌ای بهتری نسبت به سایر الگوریتم‌های آزمایش‌شده ارائه می‌دهد.

با استفاده از الگوریتم بهینه‌سازی امواج آب<sup>۷</sup> (WVO)، یک روش خوشه‌بندی ارائه شد [۳۵] که برای یافتن مراکز خوشه بهتر، در مسایل بهینه‌سازی محدودیت‌دار و یا بدون محدودیت مورد استفاده قرار گرفت.

الگوریتم‌ها راهکارهایی را پیشنهاد می‌کنند که در فرار از دام بهینه‌های محلی مؤثر هستند. عامل مهم در یافتن راه حل‌های خوب در این الگوریتم‌ها، توازن بین قابلیت‌های اکتشاف<sup>۱</sup> و بهره‌برداری<sup>۲</sup> است. قابلیت اکتشاف به جستجوی گسترده در فضای جواب اشاره دارد و قابلیت بهره‌برداری، استفاده از تجربیات به دست آمده در فرایند جستجو و تمرکز بر نواحی امیدبخش فضای جواب می‌باشد [۲۹]. بنابراین با ایجاد توازن پویا بین این دو قابلیت، جستجو به سمت محدوده‌هایی از فضای جواب سوق داده می‌شود که جواب‌های بهتری در آنها یافت شده و از طرف دیگر، موجب عدم اتلاف زمان بیشتر در بخشی از فضای جواب می‌شود که قبل از این بررسی شده و یا شامل جواب‌های نامرغوبی می‌باشد. اما این الگوریتم‌ها از آنجایی که در دسته الگوریتم‌های تقریبی هستند و تمام فضای جواب را جستجو نمی‌کنند، ممکن است به دام بهینه محلی گرفتار آیند و نتوانند از آن رهایی یابند که در این حالت جواب خوبی را بر نمی‌گردانند. یکی از الگوریتم‌هایی که در بسیاری از مسایل بهینه‌سازی استفاده شده است، الگوریتم بهینه‌سازی گرگ‌های خاکستری [۳۰] است که قابلیت بهره‌برداری خوبی دارد ولی در اکتشاف و برقراری توازن با بهره‌برداری ضعیف عمل می‌کند. در این الگوریتم، حرکت گرگ‌ها بر اساس سه گرگ برتر جمعیت صورت می‌گیرد و اگر این سه گرگ در نقطه بهینه محلی باشند کل جمعیت به سمت آن نقطه حرکت خواهند کرد.

در این تحقیق برای بهبود خوشه‌بندی داده‌ها نسخه‌ای از الگوریتم بهینه‌سازی گرگ خاکستری به نام الگوریتم بهینه‌سازی چهارگرگ<sup>۳</sup> (4GWO) ارائه می‌گردد. در الگوریتم بهینه‌سازی گرگ خاکستری، به منظور شبیه‌سازی رفتار اجتماعی گرگ‌ها، یک جمعیت تصادفی از راه حل‌ها تولید شده و اولین راه حل بهینه به نام آلفا و دومین و سومین راه حل‌های بهینه به ترتیب به نام بتا و دلتا معرفی می‌شود. سایر گرگ‌ها با استفاده از موقعیت این سه گرگ تغییر موقعیت می‌دهند، بنابراین اگر این سه گرگ در بهینه محلی گرفتار شوند سایر اعضای جمعیت را نیز به سمت این موقعیت خواهند کشاند. در الگوریتم پیشنهادی، از عملکرد دسته چهارم که گرگ‌های امگای پیشرو<sup>۴</sup> نامیده می‌شوند برای فرار از بهینه‌های محلی و بهبود قابلیت اکتشاف استفاده می‌گردد. به این منظور برای هر گرگ جمعیت، امتیازی محاسبه می‌گردد و بر اساس آن، نوع حرکت مشخص می‌شود. هرچه گرگ موقعیت بهتری داشته باشد (یعنی به جواب خوب نزدیک باشد) امتیاز آن بیشتر است. در ابتدای اجرای الگوریتم که گرگ‌ها از نقطه بهینه فاصله دارند، امتیاز کمی دارند و بنابراین بر اساس موقعیت بهترین گرگ‌های امگای پیشرو که موقعیتی تصادفی در فضای جستجو است، تغییر موقعیت می‌دهند و به اکتشاف می‌پردازند. گرگ‌ها به تدریج که از حالات اکتشاف به سمت بهره‌برداری حرکت می‌کنند، دارای امتیاز بهتری می‌گردند و بر اساس حرکت سه گرگ برتر جمعیت حرکت می‌کنند. این نحوه حرکت، باعث پیمایش بهتر فضای مسأله می‌گردد و مراکز خوشه دقیق‌تر انتخاب می‌شوند.

در ادامه و در بخش ۲ پژوهش‌های اخیر در زمینه خوشه‌بندی با الگوریتم‌های فراابتکاری شرح داده می‌شوند و پس از آن، الگوریتم بهینه‌سازی گرگ خاکستری توضیح داده می‌شود. روش پیشنهادی شامل

1. Exploration
2. Exploitation
3. 4-Gray Wolf Optimization Algorithm
4. Scout Omega Wolf

5. Software-Defined Network  
6. Distributed Denial of Service  
7. Water Wave Optimization

رنگی چهره با الگوریتم‌های ژنتیک، بهینه‌سازی ازدحام ذرات، الگوریتم بهینه‌سازی ملخ و الگوریتم جستجوی کلاغ مقایسه شد که روش پیشنهادی، عملکرد و سرعت همگرایی بهتری داشت.

از تحقیقات انجام‌شده روی خوشه‌بندی با الگوریتم‌های فراابتکاری می‌توان چنین نتیجه گرفت که این الگوریتم‌ها می‌توانند کارایی خوشه‌بندی و همچنین خوشه‌بندی‌های سنتی را بهبود دهند. تا کنون الگوریتم‌های بسیاری برای مسأله خوشه‌بندی معرفی شده‌اند، اما هنوز نیاز است تا این مسأله با استفاده از الگوریتم‌های جدیدتر که کارایی خوبی روی مسایل بهینه‌سازی با ابعاد بالا داشته‌اند و چندان نیازی به اطلاعات زمینه‌ای از مسأله ندارند، مورد بررسی و تجزیه و تحلیل قرار گیرد. از جمله این الگوریتم‌ها می‌توان به الگوریتم‌های فراابتکاری اشاره کرد که دارای این خصوصیات هستند و از میان این دسته، الگوریتم بهینه‌سازی گرگ خاکستری برای حل بسیاری از مسایل بهینه‌سازی با ابعاد بالا مورد استفاده قرار گرفته است. در ادامه به بررسی این الگوریتم پرداخته می‌شود.

## ۲-۲ الگوریتم بهینه‌سازی گرگ خاکستری

این الگوریتم از رفتار گرگ‌های خاکستری در شکار و نحوه رهبری اجتماعی آنها در طبیعت الهام گرفته شده است [۳۰] و همانند دیگر الگوریتم‌های فراابتکاری، ابتدا با یک جمعیت تصادفی از گرگ‌ها (راه حل‌های کاندیدا) آغاز می‌شود. در این الگوریتم، طبقات اجتماعی گرگ‌ها به ۴ گروه آلفا ( $\alpha$ )، بتا ( $\beta$ )، دلتا ( $\delta$ ) و امگا ( $\omega$ ) تقسیم می‌شوند. در تکرارهای مختلف الگوریتم، سه راه حل برتر هر دور به ترتیب گروه آلفا ( $\alpha$ )، بتا ( $\beta$ ) و دلتا ( $\delta$ ) نامیده می‌شوند. در این الگوریتم فرایند شکار (بهترین راه حل) توسط این سه گرگ برتر هدایت می‌شود. گرگ‌های امگا برای رسیدن به بهترین راه حل‌ها به دور این سه گرگ حلقه می‌زنند و تغییر موقعیت گرگ‌ها بر اساس روابط زیر می‌باشد

$$\bar{D} = \left| \bar{C} \cdot \bar{X}_p - \bar{X}(t) \right| \quad (۱)$$

$$\bar{X}(t+1) = \bar{X}_p - \bar{A} \cdot \bar{D} \quad (۲)$$

که  $\bar{X}_p$  بردار مکانی طعمه (شکار)،  $\bar{X}(t)$  موقعیت فعلی و  $\bar{X}(t+1)$  موقعیت بعدی را نشان می‌دهد. در روابط فوق، دو بردار  $\bar{C}$  و  $\bar{A}$  از روابط زیر محاسبه می‌شوند

$$\bar{C} = 2 \cdot \bar{r}_1 \quad (۳)$$

$$\bar{A} = 2a \cdot \bar{r}_1 - \bar{a} \quad (۴)$$

که در آن  $\bar{a}$  یک بردار کاهشی است و در طول اجرای الگوریتم به صورت خطی از ۲ به صفر می‌رسد.  $\bar{r}_1$  و  $\bar{r}_2$  بردار اعداد تصادفی در بازه [۰,۱] می‌باشند.

به منظور شبیه‌سازی ریاضیاتی رفتار گرگ‌های خاکستری در شکار، همواره موقعیت سه راه حل برتر (آلفا، بتا و دلتا) تا آخرین لحظه ذخیره شده و دیگر گرگ‌ها ( $\omega$ ) موظف هستند موقعیت خود را با توجه به موقعیت سه گرگ برتر به روز رسانی کنند. مدل ریاضیاتی به روز رسانی موقعیت گرگ‌های  $\omega$  به صورت زیر است

$$\bar{D}_\delta = \left| \bar{C}_\delta \cdot \bar{X}_\delta - \bar{X} \right|$$

$$\bar{D}_\alpha = \left| \bar{C}_\alpha \cdot \bar{X}_\alpha - \bar{X} \right| \quad (۵)$$

$$\bar{D}_\beta = \left| \bar{C}_\beta \cdot \bar{X}_\beta - \bar{X} \right|$$

محققین دیگری نیز الگوریتم‌های فراابتکاری دیگری مانند جستجوی گرانشی [۱۹]، گرگ خاکستری [۱۱] و جستجوی ارگانیسم همزیست در خوشه‌بندی مجموعه دادگان مختلف را مورد ارزیابی قرار دادند.

با پیشرفت‌های اخیر در روش‌های محاسبات مبتنی بر اینترنت، استفاده از برنامه‌های مبتنی بر ابر برای تسهیل فعالیت‌های روزانه به طور قابل توجهی در حال افزایش است. از آنجا که حجم کار ارسالی توسط کاربران برای استفاده از برنامه‌های مبتنی بر ابر از نظر معیارهای کیفیت خدمات متفاوت است، نیاز به تجزیه و تحلیل و شناسایی این حجم‌های کاری ناهمگن ابر برای منابع کارآمد به عنوان یکی از مسایل چالش‌برانگیز است. قبایی و شهیدی‌نژاد [۳۶]، روشی برای تأمین منابع کارآمد با استفاده از خوشه‌بندی مبتنی بر الگوریتم فراابتکاری برای تجزیه و تحلیل حجم کار ابر ارائه دادند. روش خوشه‌بندی پیشنهادی از ترکیبی از الگوریتم‌های ژنتیک و خوشه‌بندی فازی برای یافتن خوشه‌های مشابه با توجه به نیازهای کیفیت خدمات به کاربر استفاده می‌کند.

از الگوریتم‌های فراابتکاری برای بهبود الگوریتم‌های خوشه‌بندی سنتی مانند  $K$ -میانگین<sup>۱</sup> که اغلب در دام بهینه‌های محلی گرفتار می‌آیند و نرخ همگرایی کندی برای مجموعه داده‌های بزرگ دارند، استفاده می‌شود. با استفاده از الگوریتم ترکیبی بهینه‌سازی گرده‌افشانی گل مبتنی بر نقشه‌های آشوب و  $K$ -میانگین [۲۰] کارایی خوشه‌بندی روی مجموعه دادگان مختلف بررسی گردید که در مقایسه با سایر روش‌ها، از نظر معیارهای یکپارچگی خوشه، زمان اجرا و تعداد تکرارهای همگرایی کارایی بهتری داشت.

افراخته و بستانی با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات [۳۷]، به خوشه‌بندی داده‌های سرعت باد در نیروگاه‌های بادی پرداختند و نتایج به دست آمده را با روش‌های خوشه‌بندی فازی و  $K$ -میانگین مقایسه کردند. نتایج شبیه‌سازی نشان داد که روش ارائه‌شده همگرایی بهتری نسبت به دو روش دیگر داشت. در تحقیقی دیگر، از روش خوشه‌بندی سلسله‌مراتبی BIRCH با الگوریتم بهینه‌سازی واکنش شیمیایی جهت کشف تقلب در حوزه سلامت استفاده شد [۳۸]. نتایج نشان داد که روش پیشنهادی از سرعت و دقت بهتری در تشخیص داده‌های تقلب در حوزه سلامت نسبت به سایر الگوریتم‌های مورد مقایسه برخوردار است.

یکی دیگر از انواع خوشه‌بندی‌ها، خوشه‌بندی ظرفیت‌دار است که کاربرد وسیعی در داده‌کاوی دارد. در این مسأله، هدف افراز یک مجموعه  $n$  تایی از عناصر به  $P$  خوشه ظرفیت‌دار است به طوری که تمامی اعضای یک خوشه به نقطه‌ای که به عنوان مرکز ثقل آن خوشه تعیین می‌شود، تخصیص یابند و عدم تشابه تمامی نقاط یک خوشه از مرکز ثقل خوشه با رعایت محدودیت ظرفیت در هر خوشه حداقل گردد، به طوری که هر عنصر تنها به یک خوشه تخصیص یابد. یقینی و لسان دو روش برای این نوع خوشه‌بندی با استفاده از الگوریتم شبیه‌سازی تبریدی و الگوریتم ژنتیک ارائه کردند [۳۹]. در روش اول برای جستجوی جواب از ساختارهای مختلف همسایگی و در روش دوم از یک رویه ابتکاری جستجوی محلی استفاده شد. نتایج نشان دادند که الگوریتم ژنتیک پیشنهادی کارایی بهتری نسبت به شبیه‌سازی تبریدی در مسایل کوچک و متوسط دارد.

در پژوهشی دیگر، یک روش بخش‌بندی تصاویر رنگی چهره با استفاده از خوشه‌بند فازی بهینه‌شده با الگوریتم‌های گرگ خاکستری و نهنگ ارائه شد [۴۰]. نتایج بخش‌بندی از روش پیشنهادی در بخش‌بندی تصاویر

مستقیم در الگوریتم اصلی گرگ، یعنی نادیده گرفتن سلسله مراتب اجتماعی آن، الگوریتمی از گرگ (G-NHGW) ارائه کردند تا مشکل اکتشاف ضعیف الگوریتم گرگ خاکستری را حل کنند. همان گونه که مشاهده می شود مشکل اصلی الگوریتم بهینه سازی گرگ خاکستری، اکتشاف ضعیف و همگرایی زودرس می باشد که با بهبود آن می توان راه حل های بهتری را به دست آورد.

### ۳- روش پیشنهادی

در روش پیشنهادی، ابتدا الگوریتم چهارگرگ خاکستری توضیح داده می شود که در آن با استفاده از موقعیت گرگ های امگا پیشرو و موقعیت سایر گرگ ها، تصمیم گیری برای یافتن موقعیت بعدی هر گرگ صورت می گیرد. هدف در الگوریتم پیشنهادی، بهبود قابلیت اکتشاف و بهبود توازن بین قابلیت اکتشاف و بهره برداری است. در ادامه، نحوه نگاشت الگوریتم پیشنهادی چهارگرگ خاکستری برای یافتن مراکز خوشه در مسأله خوشه بندی بیان می گردد.

#### ۳-۱ الگوریتم بهینه سازی چهار گرگ خاکستری

در این تحقیق برای حل مشکل همگرایی زودرس و فرار از بهینه محلی الگوریتم بهینه سازی گرگ خاکستری، بهبودی از الگوریتم به نام الگوریتم بهینه سازی چهارگرگ خاکستری ارائه می گردد. ایده اصلی این الگوریتم، استفاده از دسته ای از گرگ های خاکستری امگا به نام گرگ های پیشرو در حرکت مؤثر به سمت طعمه است.

در ابتدا جمعیت گرگ ها به طور تصادفی مقداردهی اولیه می گردند و تابع برازندگی جمعیت ارزیابی می شود. بهترین موقعیت بر اساس این تابع به عنوان گرگ آلفا، بهترین راه حل دوم به عنوان گرگ بتا و بهترین راه حل سوم، گرگ دلتا در نظر گرفته می شود. موقعیت همه گرگ ها در فضای جستجو با توجه موقعیت این سه گرگ برتر به روز رسانی می گردند. بدین صورت که مقدار  $\bar{X}_1$ ،  $\bar{X}_2$  و  $\bar{X}_3$  برای هر گرگ بر اساس موقعیت گرگ های آلفا، بتا و دلتا و موقعیت فعلی گرگ ها محاسبه می شود و در نهایت موقعیت بعدی گرگ از میانگین مکانی  $\bar{X}_1$ ،  $\bar{X}_2$  و  $\bar{X}_3$  به دست می آید. در بعضی مواقع، الگوریتم در دام بهینه محلی گرفتار می آید و سه گرگ آلفا، بتا و دلتا جمعیت را به سمت بهینه محلی هدایت می کنند. برای حل این مشکل از روابط زیر، موقعیت جدید گرگ ها نسبت به گرگ آلفا محاسبه می گردد و بر اساس دور یا نزدیک بودن به گرگ آلفا، تصمیم گیرهای بعدی اتخاذ می شود

$$\overline{TP}_i = \frac{\overline{X}_{i1} + \overline{X}_{i2} + \overline{X}_{i3}}{3} \quad (8)$$

$$p_i = \frac{fit(worst) - fit(\overline{TP}_i)}{fit(worst) - fit(\bar{X}_\alpha)} \quad (9)$$

که  $i$  نشان دهنده گرگ  $i$ ام است.  $\overline{TP}_i$  موقعیت جدید موقت اول و  $fit(\overline{TP}_i)$  مقدار تابع برازندگی برای آن است.  $fit(worst)$  مقدار تابع برازندگی برای بدترین گرگ و  $fit(\bar{X}_\alpha)$  مقدار تابع برازندگی برای گرگ آلفا است.  $p_i$  میزان نزدیکی یا دوری از گرگ آلفا را نشان می دهد. مقدار  $p_i$  در بازه صفر و یک است. در صورتی که  $p_i$  برای بدترین گرگ محاسبه گردد، مقدار آن صفر و در صورتی که برای بهترین گرگ (گرگ آلفا) محاسبه شود، مقدار آن یک خواهد بود. با توجه به تعداد گرگ های

$$\begin{aligned} \bar{X}_1 &= \bar{X}_\alpha - \bar{A}_1 \cdot (\bar{D}_\alpha) \\ \bar{X}_2 &= \bar{X}_\beta - \bar{A}_2 \cdot (\bar{D}_\beta) \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{X}_3 &= \bar{X}_\delta - \bar{A}_3 \cdot (\bar{D}_\delta) \\ \overline{X(t+1)} &= \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} \end{aligned} \quad (7)$$

که در آن  $\bar{X}_\alpha$ ،  $\bar{X}_\beta$  و  $\bar{X}_\delta$  به ترتیب موقعیت گرگ های آلفا، بتا و دلتا و  $\bar{X}_1$ ،  $\bar{X}_2$  و  $\bar{X}_3$  موقعیت هر گرگ بر اساس این سه گرگ است.  $\bar{C}_1$ ،  $\bar{C}_2$  و  $\bar{C}_3$  و  $\bar{A}_1$ ،  $\bar{A}_2$  و  $\bar{A}_3$  همگی بردارهای تصادفی هستند که مانند (۳) و (۴) محاسبه می گردند و  $\overline{X(t+1)}$  موقعیت جدید گرگ ها را نشان می دهد.

تا کنون نسخه های متنوعی از الگوریتم گرگ خاکستری ارائه گردیده که در هر کدام روشی برای بهبود الگوریتم در نظر گرفته شده است. لی و همکاران [۴۱] به طور مساوی جمعیت اولیه گرگ خاکستری را در فضای مسأله توزیع کردند و از جهش گاوسی برای جلوگیری از افتادن الگوریتم در بهینه های محلی استفاده نمودند و نهایتاً یک عامل کنترل کسینوسی برای ایجاد توازن بین قابلیت اکتشاف سراسری و محلی الگوریتم در جهت بهبود سرعت همگرایی الگوریتم معرفی کردند. در پژوهشی دیگر، به منظور بهبود سرعت همگرایی، از نقشه های آشوبی در الگوریتم گرگ (CGWO) استفاده شد [۴۲]. یک الگوریتم ترکیبی با استفاده از الگوریتم بهینه ساز گرگ خاکستری و الگوریتم آتش بازی که از فرایند انفجار آتش بازی تقلید می کند، ارائه شد [۴۳] و از مزایای این دو الگوریتم برای دستیابی به بهترین نتیجه استفاده گردید. در الگوریتم پیشنهادی از قابلیت اکتشاف بهتر الگوریتم آتش بازی و قابلیت بهره برداری بهتر الگوریتم گرگ خاکستری استفاده شد. محمدرزاده و همکاران [۴۴]، یک الگوریتم بهبود یافته از گرگ خاکستری ارائه کردند که در آن ضعیف ترین گرگ ها از جمعیت حذف شده و از دیگر گرگ ها در جمعیت اولیه گنجانده می شوند. انتخاب گرگ های جایگزین به صورت تصادفی یا بر اساس تابع برازندگی خواهد بود. در این الگوریتم، موقعیت هر گرگ در هر تکرار بررسی می گردد و در صورت بهبود، این موقعیت در نظر گرفته می شود و در غیر این صورت گرگ ها در آخرین حالت مناسب باقی می ماندند.

یک الگوریتم بهبود یافته دیگر از الگوریتم گرگ (EEGWO) برای افزایش قابلیت اکتشاف گرگ خاکستری ارائه شد [۴۵]. الگوریتم پیشنهادی به منظور بهبود اکتشاف، از یک همسایه تصادفی از جمعیت برای تغییر موقعیت هر گرگ علاوه بر روابط خود الگوریتم استفاده می کند. همچنین به منظور توازن بین اکتشاف و بهره برداری، از یک پارامتر کنترل غیر خطی بهره می گیرد و پارامتر کنترلی در طول تکرارها به صورت غیر خطی تغییر می یابد. در تحقیقی دیگر، با بهبود پارامتر کنترلی  $a$  در الگوریتم گرگ (AGWO) سعی در بهبود قابلیت اکتشاف و بهره برداری شد [۴۶]. یک الگوریتم بهبود یافته دیگر (EGWO) برای تقلید از سلسله مراتب رهبری و روش شکار گروهی گرگ های خاکستری در طبیعت پیشنهاد شد [۴۷]. در این الگوریتم، محل شکار توسط گرگ های رهبر تخمین زده می شود، هر گرگ مستقیماً به سمت محل تخمینی شکار حرکت می کند و وزن گرگ آلفا در تغییر موقعیت هر گرگ بیشتر در نظر گرفته می شود. اکبری و همکاران [۴۸] با معرفی یک تغییر

1. Chaotic Grey Wolf Optimization
2. Exploration-Enhanced Grey Wolf Optimizer
3. Enhanced Grey Wolf Optimizer

ویژگی باشند، ابعاد هر گرگ به صورت شکل ۱ خواهد بود. در این شکل سه مرکز خوشه (۱ و ۱)، (۲ و ۵) و (۳ و ۴) وجود دارد که به یک گرگ اختصاص داده شده است. فاصله داده‌ها تا این مراکز محاسبه می‌گردند. بقیه گرگ‌ها نیز دارای مراکز دیگر هستند که فاصله داده‌ها تا آنها محاسبه می‌شود و از بین آنها کمترین فاصله تا مراکز خوشه محاسبه می‌گردد. تابع برازندگی که در اینجا به این منظور در نظر گرفته شده است بر اساس [۲۱] و [۴۹] می‌باشد که به صورت زیر است

$$f(D, C) = \sum_{i=1}^n \min \{ \|d_i - c_j\|^2 \}, \quad j = 1, 2, \dots, k \quad (12)$$

که هدف در آن حداقل کردن مجموع فاصله بین داده  $i$  ام  $d_i, (i = 1, 2, \dots, n)$  و مرکز خوشه  $j$  ام  $c_j, (j = 1, 2, \dots, k)$  است.  $n$  تعداد داده‌ها و  $k$  تعداد مراکز خوشه‌هاست. در رابطه فوق،  $\|d_i - c_j\|$  فاصله اقلیدسی بین داده  $i$  ام و مرکز خوشه  $j$  ام است که از رابطه زیر محاسبه می‌گردد

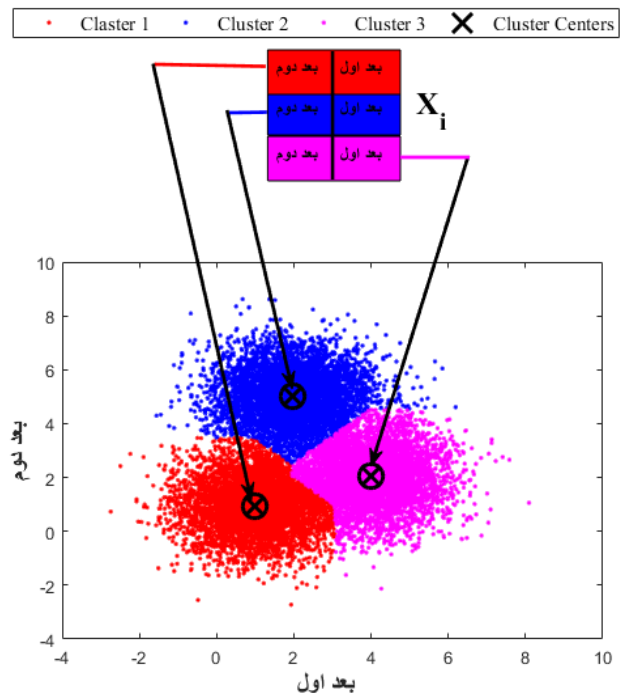
$$\text{dist}(d_i, c_j) = \sum_{p=1}^m \sqrt{(d_{ip} - c_{jp})^2} \quad (13)$$

که در آن  $m$  تعداد ابعاد داده‌هاست.

در واقع هدف اصلی، اختصاص داده‌ها به مراکز خوشه‌ها است به نحوی که فواصل داده‌های درون خوشه با مرکز آن کمینه گردند. در روش پیشنهادی، پارامترهای الگوریتم مقداردهی می‌شوند و بر اساس تعداد ویژگی‌های مجموعه دادگان و تعداد مراکز خوشه‌ها، ابعاد هر گرگ مشخص می‌شود. سپس مراکز خوشه به صورت تصادفی مقداردهی می‌گردند. موقعیت هر گرگ به عنوان راه حلی برای مراکز خوشه در نظر گرفته می‌شود. الگوریتم اجرا می‌شود و فاصله نمونه‌های مجموعه دادگان تا مراکز خوشه‌ها محاسبه می‌گردد. داده‌ها در خوشه‌هایی قرار می‌گیرند که حداقل فاصله تا مراکز خوشه‌ها را داشته باشند. گرگ‌ها راه حل‌های جدید را پیدا کرده و با توجه به تابع برازندگی آنها، سه گرگ برتر انتخاب می‌شوند. بر اساس موقعیت هر گرگ، تصمیم‌گیری برای تعیین موقعیت بعدی انجام می‌گردد. در صورتی که موقعیت گرگ از گرگ آلفا دور باشد، حرکت بر اساس موقعیت بهترین گرگ امگای پیشرو و سه گرگ برتر صورت می‌گیرد و در غیر این صورت فقط بر اساس موقعیت سه گرگ برتر انجام می‌شود. این روال تا رسیدن به شرط خاتمه الگوریتم ادامه پیدا می‌کند و در پایان موقعیت گرگ آلفا به عنوان بهترین مراکز خوشه برمی‌گردد. شبه‌کد الگوریتم پیشنهادی در شکل ۲ نشان داده شده است.

### ۳-۳ مثالی از روش خوشه‌بندی با استفاده از الگوریتم پیشنهادی

در جدول ۱ موقعیت‌های اعضای جمعیت در الگوریتم بهینه‌سازی چهارگرگ در تکرارهای مختلف نشان داده شده است. در ابتدا اعضای جمعیت مقداردهی اولیه می‌شوند و ابعاد هر گرگ به تعداد مراکز خوشه در تعداد ویژگی‌های مجموعه دادگان است. مشخصات مجموعه داده در ستون اول جدول ۲ نمایش داده شده است. در این مثال مجموعه داده دارای ۲ ویژگی (Feature ۱, Feature ۲) است و برچسب داده‌های آن (Label) و تعداد نمونه‌های آن ۱۰ می‌باشد. در مرحله مقداردهی اولیه (سطر اول جدول ۱) ابتدا همه گرگ‌ها به صورت تصادفی مقداردهی اولیه می‌شوند (Initialization) و از بین آنها بهترین گرگ که دارای بهترین تابع برازندگی است به عنوان گرگ آلفا انتخاب می‌گردد. در جدول ۱ در ردیف اول، گرگ ۲ (GW۲) که دارای بهترین تابع برازندگی (۱۲,۳۱۱۹)



شکل ۱: ابعاد و ساختار هر گرگ در مسأله خوشه‌بندی.

امگای پیشرو، این دسته به صورت تصادفی در فضای مسأله، مقداردهی می‌گردد و بهترین موقعیت گرگ‌های امگای پیشرو ( $X_{best_{sw}}$ ) انتخاب و موقعیت موقت دوم برای گرگ  $i$  ام از رابطه زیر محاسبه می‌گردد

$$\overline{TP}_{\nu_i} = \overline{TP}_{\nu_i} + \text{rand}(-1, 1)(\overline{TP}_{\nu_i} - X_{best_{sw}}) \quad (10)$$

که در آن  $\text{rand}(-1, 1)$  یک عدد تصادفی بین ۱ و -۱ است. حال بر اساس  $p_i$ ، تصمیم‌گیری برای موقعیت جدید گرگ  $i$  ام به صورت زیر انجام می‌گیرد

$$\overline{X}_i(t+1) = \begin{cases} \overline{TP}_{\nu_i} & \text{if } p_i \geq 0.5 \\ \overline{TP}_{\nu_i} & \text{otherwise} \end{cases} \quad (11)$$

اگر  $p_i \leq 0.5$  باشد، نشان‌دهنده آن است که موقعیت موقت اول از گرگ آلفا فاصله دارد و موقعیت جدید برابر موقعیت موقت دوم می‌شود و در غیر این صورت، موقعیت جدید برابر موقعیت موقت اول می‌گردد.

در ابتدای الگوریتم باید اکتشاف فضاهای جدید صورت گیرد، بنابراین موقعیت جدید گرگ بر اساس موقعیت بهترین گرگ امگای پیشرو (که به صورت تصادفی مقدار گرفته است) محاسبه می‌گردد. به تدریج در اواسط اجرای الگوریتم، بایستی گرگ‌ها از حالت اکتشاف به سمت بهره‌برداری حرکت کنند. از این رو موقعیت جدید بر اساس موقعیت موقت اول که نزدیک‌ترین موقعیت یافته‌شده تا کنون (گرگ آلفا) است، محاسبه می‌گردد. شرط خاتمه در این الگوریتم بر اساس شرایط مسأله مشخص می‌گردد که می‌تواند رسیدن به تعداد تکرار مشخص یا خطای مشخص شده در مسأله باشد.

### ۲-۳ الگوریتم پیشنهادی چهارگرگ خاکستری در مسأله خوشه‌بندی

به منظور تطبیق الگوریتم پیشنهادی با خوشه‌بندی، موقعیت هر گرگ به عنوان راه حلی برای مراکز خوشه در نظر گرفته می‌شود، یعنی موقعیت هر گرگ شامل مراکز خوشه‌ها می‌باشد. به عنوان مثال اگر ۳ مرکز خوشه در نظر گرفته شود و داده‌هایی که قرار است خوشه‌بندی گردند دارای ۲

سیستم عامل bit ۶۴ و در محیط متلب ۲۰۱۵ پیاده‌سازی شده‌اند. در ابتدا مقدار پارامترها با استفاده از نتایج آزمایش‌ها به دست می‌آید و سپس روش پیشنهادی بر روی مجموعه دادگان UCI [۵۰] که در [۲۱]، [۲۵] و [۴۹] استفاده شده است، با الگوریتم‌هایی مانند الگوریتم بهینه‌ساز گرگ خاکستری، بهینه‌سازی ازدحام ذرات، زنبور عسل مصنوعی، ارگانسیم‌های همزیست و سالپ مقایسه می‌شود. مجموعه دادگان UCI مورد استفاده و ویژگی‌های آنها در جدول ۳ نشان داده شده است. این مجموعه دادگان دارای تعداد ویژگی کم، متوسط و زیاد می‌باشند. همچنین در تعداد مراکز خوشه و تعداد نمونه نیز دارای این تنوع می‌باشند که قرارگرفتن صحیح داده‌ها در خوشه‌های مناسب در مجموعه دادگانی با تعداد مراکز بالا نیز از چالش‌های خوشه‌بندی است. علاوه بر آن در این جدول، مجموعه دادگانی با داده‌های نامتوازن نیز در نظر گرفته شده که خوشه‌بندی آنها مشکل می‌باشد و از جمله آنها می‌توان به مجموعه دادگان Zoo و Ecoli اشاره کرد.

شبیه‌سازی و نتایج پیاده‌سازی الگوریتم پیشنهادی ۴GWO-C بر اساس تعداد اعضای جمعیت، تعداد ابعاد مسأله و حداکثر تعداد تکرار به عنوان متغیرهای مستقل و درصد خطا یا طبقه‌بندی نادرست (MCR)، دقت، معیار  $F^3$  و مجموع فاصله اعضا تا مراکز خوشه (تابع برازندگی) به عنوان متغیر وابسته برای مشخص نمودن پارامترهای الگوریتم مورد ارزیابی قرار می‌گیرد که این نتایج در جدول ۴ آمده است. سپس با استفاده از معیارهای ارزیابی، کارایی الگوریتم‌ها با یکدیگر مقایسه می‌شوند. معیارهای ارزیابی به صورت زیر محاسبه می‌گردند [۲۵] و [۴۹]

$$P(i, j) = \frac{n_{i,j}}{n_j} \quad (14)$$

$$R(i, j) = \frac{n_{i,j}}{n_i} \quad (15)$$

$$f(i, j) = 2 \times \frac{P(i, j) \times R(i, j)}{P(i, j) + R(i, j)} \quad (16)$$

که  $P(i, j)$  و  $R(i, j)$  به ترتیب دقت و فراخوانی،  $n_{i,j}$  تعداد اعضای درست کلاس  $i$  در خوشه  $j$ ،  $n_j$  تعداد همه اعضای خوشه  $j$  و  $n_i$  تعداد همه اعضای کلاس  $i$  است. معیار  $F$  برای همه خوشه‌ها از رابطه زیر محاسبه می‌گردد

$$F - \text{measure} = \sum_j \frac{n_j}{n} \max_i f(i, j) \quad (17)$$

که در آن  $n$  تعداد اعضای همه خوشه‌هاست. همچنین درصد طبقه‌بندی نادرست از رابطه زیر بر حسب درصد تعداد طبقه‌بندی‌های نادرست به تعداد کل طبقه‌بندی‌ها محاسبه می‌گردد که در یک خوشه‌بندی ایده‌آل این مقدار صفر است

$$MCR = \frac{\text{Number of Incorrect Classifications}}{\text{Total Number of Classifications}} \times 100 \quad (18)$$

نتایج ارزیابی الگوریتم ۴GWO-C در مسأله خوشه‌بندی روی مجموعه دادگان UCI در مقایسه با الگوریتم‌های دیگر، حاصل از ۱۰ بار اجرای مستقل و میانگین‌گیری روی نتایج، در جدول ۵ آمده است. همان گونه

1. Misclassification Rate
2. Precision
3. F-Measure

#### Algorithm: 4GWO algorithm for Clustering (4GWO-C)

1. Function 4GWO-C (problem)
2. **Input:** Population\_size, ScoutOmega\_size, dataset
3. **Output:** Best solution ( $\bar{X}_\alpha$  : Cluster centers)
4. Set parameters
5. Compute the dimension of each wolf
6. Initialize  $X_i, i=1,2,\dots,Population\_size$
7. **Repeat**
8. Evaluate  $f(D, C) = \sum_{i=1}^n \min_j \|d_i - c_j\|, j=1,2,\dots,k$
9. Select  $\bar{X}_\alpha, \bar{X}_\beta, \bar{X}_\delta$  and the worst wolf
10. **for**  $i = 1$  **to**  $Population\_size$  **do**
11. Compute  $\bar{X}_{i1}, \bar{X}_{i2}, \bar{X}_{i3}$
12. Compute  $\overline{TP1}_i = \frac{\bar{X}_{i1} + \bar{X}_{i2} + \bar{X}_{i3}}{3}$
13. Compute  $p_i = \frac{fit(worst) - fit(\overline{TP1}_i)}{fit(worst) - fit(\bar{X}_\alpha)}$
14. Initialize  $\bar{X}_{soj}, j=1,2,\dots,Scout\_Omega\_size$
15. Select the best of  $\bar{X}_{soj}$  as  $\bar{X}_{best\_soj}$
16. Compute  $\overline{TP2}_i = \overline{TP1}_i + rand(-1,1)(\overline{TP1}_i - \bar{X}_{best\_soj})$
17. Compute  $\bar{X}_i = \begin{cases} \overline{TP1}_i & \text{if } p_i \geq 0.5 \\ \overline{TP2}_i & \text{otherwise} \end{cases}$
18. **End for**
19. **Until stop criteria is met**
20. Best solution =  $\bar{X}_\alpha$
21. Return Best Solution

شکل ۲: شبه‌کد الگوریتم بهینه‌سازی چهارگرگ برای خوشه‌بندی داده‌ها.

است به عنوان گرگ آلفا انتخاب می‌شود و همان طور که در جدول ۲ و شکل ۳ دیده می‌شود، مرکز خوشه ۱ در این مرحله فقط یک عضو دارد. بر اساس حرکت گرگ‌ها در اولین تکرار (Iteration = ۱) مراکز خوشه‌ها تغییر می‌کند ولی باز همان گرگ شماره ۲ بهترین نتیجه را دارد که تابع برازندگی آن از همه بهتر است. مرکز خوشه ۱ در این مرحله دارای دو عضو می‌گردد، همان طور که در جدول ۲ و شکل ۳ نشان داده شده است. همان گونه که در جدول‌های ۱ و ۲ و شکل ۳ دیده می‌شود، در تکرارهای ۲، ۳، ۴، گرگ ۱ (GW1) دارای بهترین مراکز خوشه است و پس از پنج تکرار (Iteration = ۵) خوشه ۱ دارای ۵ عضو و خوشه ۲ نیز دارای ۵ عضو می‌گردد و بهترین مراکز خوشه توسط گرگ ۲ به دست می‌آید که کمترین مقدار تابع برازندگی را دارد.

در شکل ۴، نمودار تابع برازندگی به دست آمده توسط گرگ آلفا در تکرارهای متوالی رسم شده است. تابع برازندگی بر اساس (۱۲) محاسبه می‌گردد. همان طور که در شکل دیده می‌شود در تکرارهای متوالی، مجموع فواصل عناصر هر خوشه تا مرکز آن خوشه کم می‌گردد تا در تکرار پنجم (Iteration = ۵)، مجموع فواصل عناصر هر خوشه تا مراکز به حداقل می‌رسد. نتایج جدول ۲ و شکل‌های ۳ و ۴ نشان می‌دهد که در این مجموعه داده کوچک پس از پنج تکرار، مراکز خوشه‌ها به درستی انتخاب می‌شوند و داده‌ها در خوشه‌های اصلی قرار می‌گیرند.

#### ۴- ارزیابی الگوریتم پیشنهادی و ارائه نتایج

در این بخش به ارزیابی الگوریتم پیشنهادی پرداخته می‌شود. الگوریتم‌ها بر روی سیستمی با Intel core i5، حافظه اصلی GB ۸ و

جدول ۱: مقادیر موقعیت‌های اعضای جمعیت در الگوریتم بهینه‌سازی چهارگرگ در تکرارهای مختلف.

Iteration	GW <sub>1</sub>		GW <sub>2</sub>		GW <sub>3</sub>		GW <sub>4</sub>		$\bar{X}_\alpha$	Best result	
Initialization	-۰٫۷۳۸۳	۰٫۴۶۴۰	۱٫۵۸۹۴	۰٫۰۰۳۰	۱٫۶۳۴۳	۰٫۵۳۶۸	۱٫۶۸۷۴	۰٫۲۰۳۰	۱٫۵۸۹۴	۰٫۰۰۳۰	۱۲٫۳۱۱۹
	۰٫۱۰۰۴	۰٫۱۵۶۲	۱٫۰۶۵۱	۰٫۷۵۳۴	-۰٫۶۷۵۷	۱٫۷۳۴۴	۰٫۱۶۳۱	۱٫۰۵۳۴	۱٫۰۶۵۱	۰٫۷۵۳۴	
۱	-۰٫۶۷۵۵	۰٫۱۹۴۷	۱٫۳۳۲۴	۰٫۲۲۶۴	۱٫۳۷۷۳	۰٫۳۰۹۱	۰٫۳۵۲۴	۱٫۷۴۶۴	۱٫۳۳۲۴	۰٫۲۲۶۴	۱۱٫۸۹۳
	۰٫۲۶۷۴	۰٫۴۲۵۵	۰٫۸۰۸۱	۰٫۶۸۱۹	-۰٫۴۱۸۷	۱٫۴۶۵۰	۱٫۸۰۹۱	۰٫۶۲۲۹	۰٫۸۰۸۱	۰٫۶۸۱۹	
۲	-۰٫۴۱۸۵	-۰٫۰۰۰۹	۱٫۰۷۵۴	۰٫۱۵۱۴	۱٫۱۲۰۳	۰٫۰۲۹۸	۰٫۴۱۹۵	۰٫۹۰۰۹	-۰٫۴۱۸۵	-۰٫۰۰۰۹	۱۱٫۰۰۷۵
	۰٫۵۲۴۴	۰٫۶۹۴۹	۰٫۵۵۱۱	۰٫۶۳۰۱	-۰٫۱۶۱۸	۱٫۱۹۵۷	-۰٫۵۷۴۴	۱٫۰۰۴۹	۰٫۵۲۴۴	۰٫۶۹۴۹	
۳	-۰٫۲۳۱۰	-۰٫۱۴۳۶	۰٫۸۱۸۵	-۰٫۱۱۸۰	۰٫۸۶۳۴	-۰٫۱۵۸۶	-۱٫۲۳۱۰	-۰٫۱۴۳۶	-۰٫۲۳۱۰	-۰٫۱۴۳۶	۹٫۹۸۹۵
	۰٫۷۱۱۹	۰٫۸۹۱۵	۰٫۶۰۳۶	۰٫۸۴۳۶	۰٫۰۹۵۲	۰٫۹۲۶۳	۰٫۹۲۲۹	۰٫۰۹۱۵	۰٫۷۱۱۹	۰٫۸۹۱۵	
۴	-۰٫۰۹۴۱	-۰٫۱۵۸۶	۰٫۵۶۱۵	-۰٫۱۵۸۶	۰٫۶۰۶۴	۰٫۰۲۶۱	-۱٫۰۹۴۱	-۰٫۱۵۸۶	-۰٫۰۹۴۱	-۰٫۱۵۸۶	۹٫۲۶۴۷
	۰٫۸۴۸۸	۱٫۰۳۵۰	۰٫۸۶۰۶	۱٫۰۹۶۵	۰٫۳۵۲۱	۰٫۷۶۴۲	۰٫۸۴۸۸	۱٫۰۳۵۰	۰٫۸۴۸۸	۱٫۰۳۵۰	
۵	۰٫۰۰۵۸	-۰٫۰۸۲۶	۰٫۳۰۴۶	۰٫۱۱۰۷	۰٫۳۴۹۵	۰٫۱۷۰۵	۱٫۳۰۴۲	۰٫۶۱۱۷	۰٫۳۰۴۶	۰٫۱۱۰۷	۷٫۹۳۴۲
	-۰٫۹۴۸۷	۱٫۱۳۹۷	۱٫۱۱۷۶	۱٫۳۳۱۶	۰٫۶۰۹۱	۱٫۰۳۳۶	۰٫۲۱۷۶	-۱٫۳۴۲۶	۱٫۱۱۷۶	۱٫۳۳۱۶	

جدول ۲: مراکز خوشه‌های به دست آمده در تکرارهای متوالی.

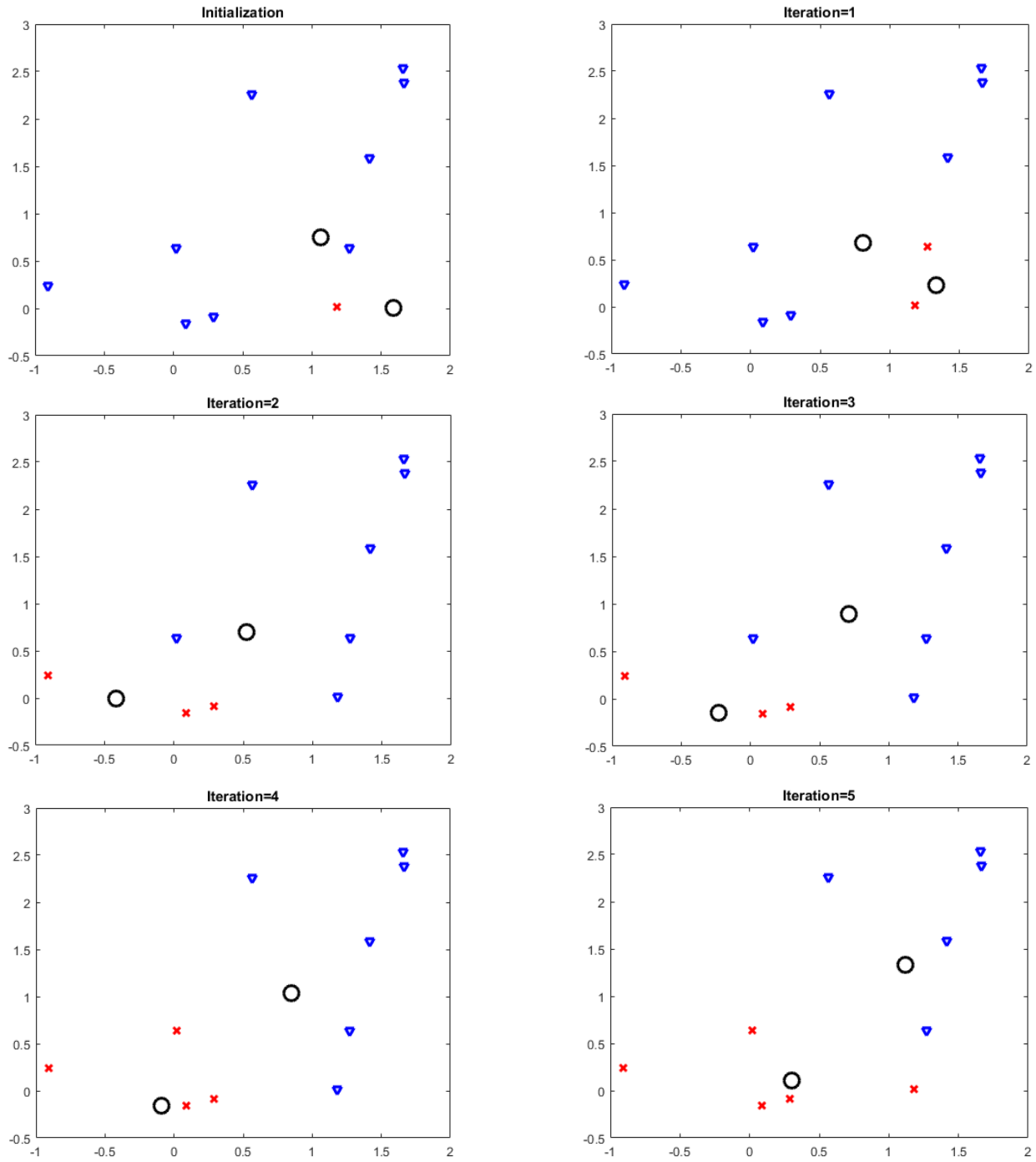
Dataset	Step		Data Clustering						
	Feature <sub>1</sub>	Feature <sub>2</sub>	Label	Initialization	Iteration = ۱	Iteration = ۲	Iteration = ۳	Iteration = ۴	Iteration = ۵
	۰٫۲۹۱	-۰٫۰۹۱	۱	۲	۲	۱	۱	۱	۱
	۱٫۱۸۲	۰٫۰۲۱	۱	۱	۱	۲	۲	۲	۱
	-۰٫۹۰۵	۰٫۲۴۲	۱	۲	۲	۱	۱	۱	۱
	۰٫۰۲۳	۰٫۶۴۲	۱	۲	۲	۲	۲	۱	۱
	۰٫۰۸۶	-۰٫۱۵۹	۱	۲	۲	۱	۱	۱	۱
	۱٫۶۵۹	۲٫۵۳۵	۲	۲	۲	۲	۲	۲	۲
	۱٫۲۷۱	۰٫۶۳۵	۲	۲	۱	۲	۲	۲	۲
	۱٫۶۶۵	۲٫۳۷۹	۲	۲	۲	۲	۲	۲	۲
	۰٫۵۳۶	۲٫۲۶۲	۲	۲	۲	۲	۲	۲	۲
	۱٫۴۱۴	۱٫۵۸۳	۲	۲	۲	۲	۲	۲	۲

جدول ۳: مجموعه دادگان مورد استفاده.

No.	Dataset	Number of features	Number of cluster	Number of instances
۱	Iris	۴	۳	۱۵۰
۲	Breast Cancer Wisconsin (Original)	۹	۲	۶۸۳
۳	Balance scale	۴	۳	۶۲۵
۴	Seeds	۷	۳	۲۱۰
۵	Statlog (Heart)	۱۳	۲	۲۷۰
۶	Wine	۱۳	۳	۱۷۸
۷	Glass	۹	۷	۲۱۴
۸	Zoo	۱۶	۷	۱۰۱
۹	Solar	۶۰	۲	۲۰۸
۱۰	Tae	۵	۳	۱۵۱
۱۱	PenglungEW	۳۲۵	۷	۷۳
۱۲	Parkinson Disease	۲۲	۲	۱۹۵
۱۳	Ecoli	۷	۸	۳۳۶
۱۴	Vehicle	۱۸	۴	۸۴۶

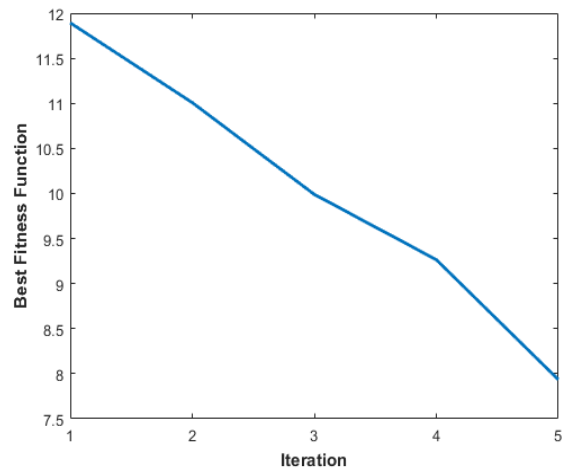
دارای ویژگی‌ها، مراکز خوشه و تعداد متفاوتی نمونه هستند. همچنین برخی از این مجموعه دادگان، داده‌های نامتوازن هستند که نمونه‌های یک یا چند کلاس، چند برابر کلاس‌های دیگر است که کار خوشه‌بندی را مشکل می‌سازد. از طرفی تعداد زیاد خوشه‌ها، چالش دیگری است که در جهت عکس بهبود معیارهای ارزیابی است.

که مشاهده می‌شود، الگوریتم‌ها بر اساس درصد طبقه‌بندی نادرست (خطا)، انحراف معیار، دقت و معیار  $F$  حاصل از نتایج خوشه‌بندی با یکدیگر مقایسه شده‌اند. تعداد جمعیت اولیه ۵۰ و حداکثر تکرار ۱۵۰ در نظر گرفته شده و هر کدام از الگوریتم‌ها ۱۰ بار اجرا شده‌اند و میانگین نتایج آنها در جدول قرار گرفته است. مجموعه دادگان در نظر گرفته شده،



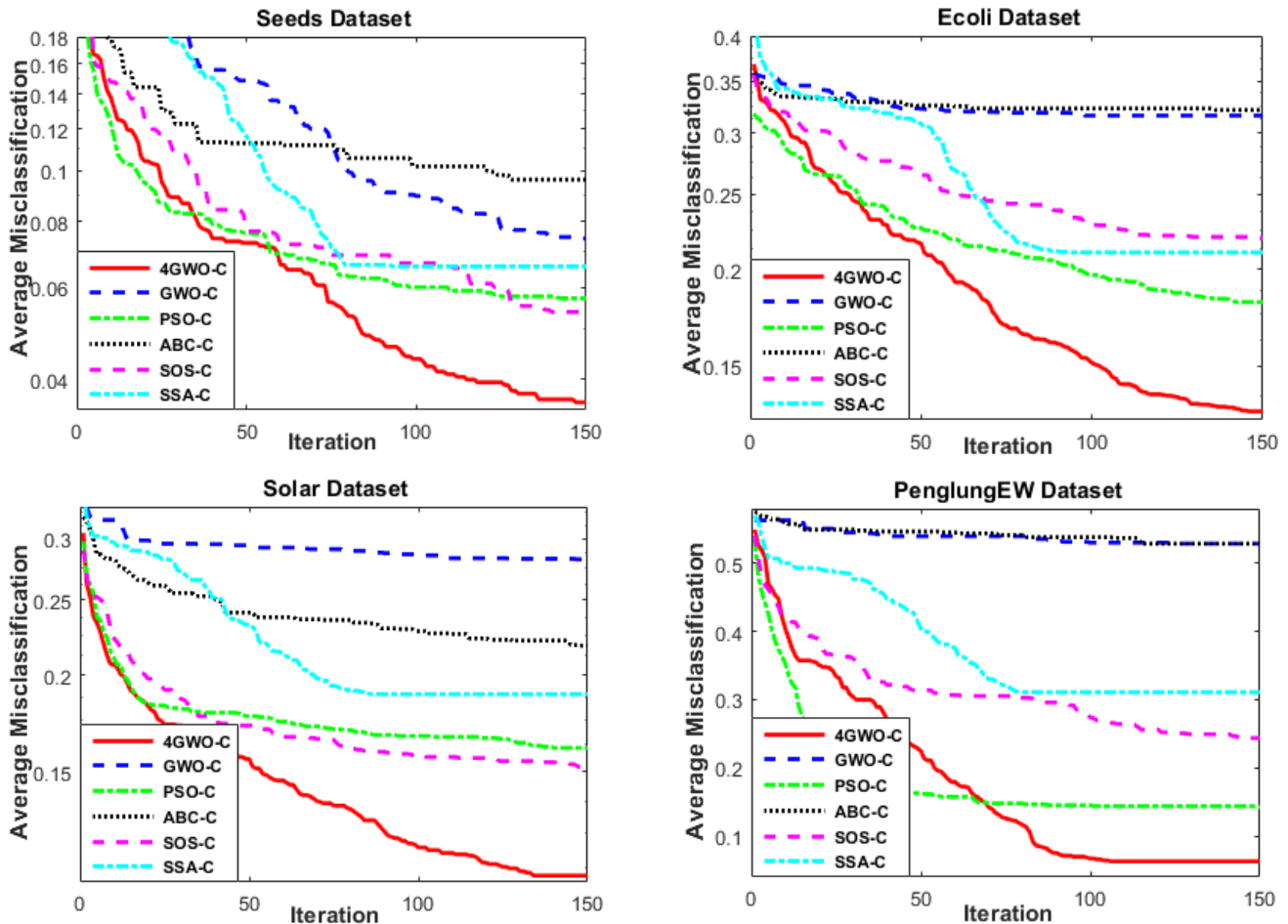
شکل ۳: مراحل تغییر مراکز خوشه‌ها.

الگوریتم‌های مورد مقایسه، الگوریتم پیشنهادی  $\text{GWO-C}$ ، الگوریتم گرگ خاکستری، الگوریتم زنبور عسل مصنوعی، الگوریتم ارگانیسم‌های همزیست، الگوریتم سالپ و الگوریتم ازدحام ذرات می‌باشند. همان طور که در جدول ۵ قابل مشاهده است در بین این الگوریتم‌ها، نتایج الگوریتم پیشنهادی بسیار قابل توجه است و راه حل‌های بهتری را ارائه داده است. در مجموعه دادگان Iris و Breast Cancer نتایج الگوریتم‌ها مشابه یکدیگر می‌باشد، هرچند که الگوریتم پیشنهادی و SOS نتایج بهتری را ارائه می‌دهند. الگوریتم گرگ خاکستری در تمام مجموعه دادگان به غیر از دو مجموعه دادگان Balance scale و Seeds، نتایج بسیار ضعیفی از خود ارائه می‌دهد. دلیل این امر آن است که در الگوریتم گرگ خاکستری، جمعیت دنباله‌رو سه عضو بهتر هستند و چنانچه این سه عضو به بهینه محلی همگرا گردند، تمام جمعیت به سمت آن بهینه کشیده



شکل ۴: نمودار تغییرات تابع برازندگی گرگ آلفا (بهترین راه حل).





شکل ۵: رفتار الگوریتم‌ها بر اساس خطای دسته‌بندی روی مجموعه دادگان Seeds, Ecoli, Solar, و PenglungEW.

مختلفی مثل ترکیب خوشه‌بندی و طبقه‌بندی برای طبقه‌بندی آنها استفاده می‌شود. به همین دلیل تفاوت قابل توجهی بین درصد طبقه‌بندی نادرست (خطا) و دقت در آنها دیده می‌شود. به عنوان مثال، مجموعه دادگان Ecoli دارای ۸ کلاس با ۳۳۶ نمونه است. از بین این نمونه‌ها دو نمونه متعلق به کلاس ۳، ۲ نمونه متعلق به کلاس ۴ و ۵ نمونه متعلق به کلاس ۷ هستند. بیشترین نمونه متعلق به کلاس ۱ و کلاس ۲ است. در این مجموعه دادگان، عملکرد الگوریتم گرگ خاکستری و الگوریتم زنبور عسل مصنوعی بسیار ضعیف است.

در شکل ۵، رفتار الگوریتم‌ها بر اساس خطای دسته‌بندی روی مجموعه دادگان Seeds, Ecoli, Solar, و PenglungEW نشان داده شده است. با توجه به این شکل، الگوریتم بهینه‌سازی گرگ خاکستری و الگوریتم زنبور عسل در همان تکرارهای اولیه به دام بهینه محلی گرفتار می‌شوند و توانایی فرار از آن را ندارند. اما روش پیشنهادی با اکتشاف خوب در تکرارهای اولیه و برقراری توازن با بهره‌برداری از اواسط اجرای الگوریتم، مراکز خوشه بهتری را انتخاب می‌کند.

کارایی بهتر روش پیشنهادی در به دست آوردن پاسخ‌های خوب، به دلیل عملکرد الگوریتم در برقراری توازن بین اکتشاف و بهره‌برداری است. با محاسبه  $p_i$  برای هر گرگ در الگوریتم پیشنهادی، موقعیت هر گرگ نسبت به گرگ آلفا مشخص می‌گردد. در ابتدای اجرای الگوریتم بایستی اکتشاف صورت گیرد و سپس از اواسط اجرا از اکتشاف به سمت بهره‌برداری حرکت کرد. در ابتدا که گرگ‌ها از یکدیگر فاصله دارند،  $p_i$  آنها کم است و این نشان می‌دهد که از بهترین راه حل فاصله دارند و در نتیجه به اکتشاف می‌پردازند. به تدریج که به اواسط اجرای الگوریتم

جدول ۴: تعیین پارامترهای الگوریتم 4GWO-C.

Variable	Metric	MCR	STD	Precision	F-Measure
Population-Size = ۵۰					
Iter = ۱۰۰		۳,۰۶۹٪	۰,۰۱۳۶	۹۰,۶۴۲٪	۹۱,۰۶۹٪
Iter = ۱۵۰		۱,۸۸۱٪	۰,۰۱۵۸	۹۶,۱۳۷٪	۹۵,۴۴۰٪
Iter = ۲۰۰		۲,۹۷۰٪	۰,۰۱۷۵	۸۹,۸۸۳٪	۹۰,۵۸۰٪
Max-Iteration = ۱۵۰					
Population-Size = ۲۰		۳,۹۶۰٪	۰,۰۲۵۶	۸۵,۱۷۷٪	۸۶,۷۵۹٪
Population-Size = ۳۰		۲,۹۷۰٪	۰,۰۱۹۲	۹۳,۴۲۷٪	۹۲,۶۹۳٪
Population-Size = ۵۰		۱,۸۸۱٪	۰,۰۱۵۸	۹۶,۱۳۷٪	۹۵,۴۴۰٪
Population-Size = ۵۰ و Max-Iteration = ۱۵۰					
ScoutOmega_size = ۵٪		۳,۴۶۵٪	۰,۰۲۳	۹۲,۲۷۵٪	۹۲,۰۳۳٪
ScoutOmega_size = ۱۰٪		۱,۸۸۱٪	۰,۰۱۵۸	۹۶,۱۳۷٪	۹۵,۴۴۰٪
ScoutOmega_size = ۲۰٪		۳,۴۶۵٪	۰,۰۲۲	۹۰,۲۴۵٪	۹۰,۷۸۳٪

می‌شوند و امکان فرار از بهینه محلی را ندارند. یعنی اگرچه الگوریتم گرگ خاکستری دارای قابلیت بهره‌برداری خوبی است اما قابلیت اکتشاف ضعیفی دارد. در مجموعه دادگان Glass، الگوریتم‌های گرگ خاکستری و زنبور عسل مصنوعی، نتایج بسیار ضعیفی مخصوصاً در معیار دقت و معیار  $F$  نشان می‌دهد. در مجموعه دادگان Zoo, Solar, و PenglungEW، Ecoli و Vehicle تفاوت نتایج الگوریتم پیشنهادی با بقیه الگوریتم‌ها بسیار قابل توجه است. این مجموعه دادگان در دسته مجموعه دادگان نامتوازن می‌باشند که برای طبقه‌بندی بسیار مشکل هستند و از روش‌های

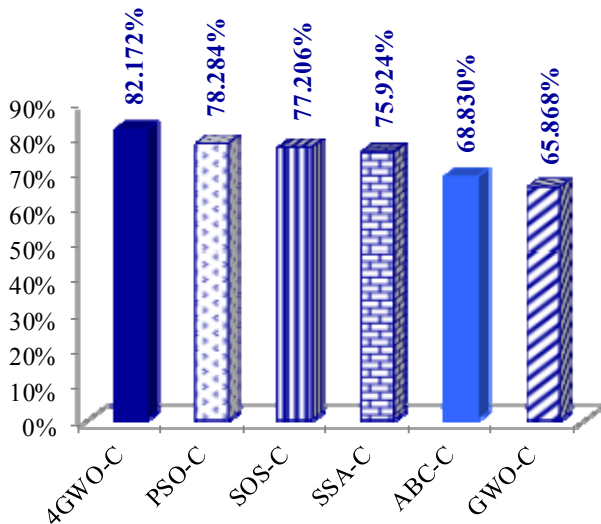
جدول ۵: نتایج حاصل از اجرای الگوریتم روی مجموعه دادگان.

Metric	Dataset	MCR	STD	Precision	F-Measure	Dataset	MCR	STD	Precision	F-Measure
Algorithm										
Iris	۴GWO-C	۰.۹۳۳٪	۰.۰۰۳۴	۹۹.۰۸۰٪	۹۹.۰۷۳٪	Zoo	۱.۸۸۱٪	۰.۰۱۵۸	۹۶.۱۳۷٪	۹۵.۴۴۰٪
	GWO-C	۱.۱۳۳٪	۰.۰۰۳۲	۹۸.۸۸۶٪	۹۸.۸۷۶٪		۲۱.۶۸۳٪	۰.۰۳۴۸	۴۸.۰۰۵٪	۵۱.۱۷۱٪
	PSO-C	۱.۶۶۷٪	۰.۰۰۹۶	۹۸.۴۰۰٪	۹۸.۳۶۶٪		۸.۶۱۴٪	۰.۰۴۶۷	۸۲.۱۵۹٪	۸۲.۰۹۴٪
	ABC-C	۲.۴۰۰٪	۰.۰۰۵۶	۹۷.۷۲۴٪	۹۷.۶۶۲٪		۱۹.۰۱۰٪	۰.۰۱۶۰	۵۷.۸۷۲٪	۵۹.۷۹۷٪
	SOS-C	۱.۱۳۳٪	۰.۰۰۵۵	۹۸.۸۹۱٪	۹۸.۸۷۹٪		۷.۶۲۴٪	۰.۰۲۳۸	۸۵.۴۲۴٪	۸۳.۵۹۶٪
	SSA-C	۱.۶۰۰٪	۰.۰۰۷۸	۹۸.۴۴۳٪	۹۸.۴۲۲٪		۱۰.۴۹۵٪	۰.۰۴۷۶	۷۹.۹۱۰٪	۷۹.۹۷۰٪
	۴GWO-C	۱.۹۴۷٪	۰.۰۰۱۲	۹۷.۵۴۵٪	۹۷.۸۹۲٪		۱۱.۰۱۰٪	۰.۰۱۱۹	۸۹.۲۸۵٪	۸۹.۰۵۸٪
	GWO-C	۲.۱۲۳٪	۰.۰۰۲۱	۹۷.۲۵۳٪	۹۷.۷۲۱٪		۲۸.۲۲۱٪	۰.۰۲۴۹	۷۲.۲۰۶٪	۷۱.۸۷۳٪
	PSO-C	۲.۰۵۰٪	۰.۰۰۱۰	۹۷.۴۶۱٪	۹۷.۷۷۷٪		۱۶.۱۰۶٪	۰.۰۲۰۰	۸۴.۳۹۱٪	۸۴.۱۹۱٪
	ABC-C	۲.۱۸۲٪	۰.۰۰۱۳	۹۷.۲۵۱٪	۹۷.۶۴۵٪		۲۱.۸۲۷٪	۰.۰۱۳۶	۷۸.۷۷۷٪	۷۸.۳۲۴٪
	SOS-C	۱.۹۴۷٪	۰.۰۰۰۷	۹۷.۵۰۲٪	۹۷.۸۹۹٪		۱۵.۱۴۴٪	۰.۰۱۴۹	۸۵.۳۶۸٪	۸۴.۹۵۱٪
	SSA-C	۲.۲۲۵٪	۰.۰۰۲۷	۹۷.۲۵۵٪	۹۷.۵۹۱٪		۱۸.۸۹۴٪	۰.۰۳۱۰	۸۱.۲۹۴٪	۸۱.۲۰۳٪
Breast Cancer Wisconsin	۴GWO-C	۱۴.۷۶۸٪	۰.۰۱۴۴	۷۰.۸۰۴٪	۷۱.۲۵۰٪	Solar	۳۹.۲۷۲٪	۰.۰۲۱۴	۶۳.۴۷۶٪	۶۲.۲۶۹٪
	GWO-C	۱۶.۶۵۶٪	۰.۰۱۷۱	۶۹.۲۱۶٪	۶۹.۹۱۰٪		۴۳.۵۱۰٪	۰.۰۵۰۲	۵۷.۳۶۷٪	۵۶.۸۶۴٪
	PSO-C	۱۹.۵۸۴٪	۰.۰۲۵۳	۶۷.۱۷۱٪	۶۷.۶۰۷٪		۳۹.۶۰۳٪	۰.۰۲۳۱	۶۳.۰۴۶٪	۶۱.۸۵۷٪
	ABC-C	۲۴.۲۷۲٪	۰.۰۲۳۱	۶۵.۷۲۳٪	۶۵.۸۲۶٪		۴۱.۱۹۲٪	۰.۰۱۶۲	۶۰.۶۴۹٪	۵۹.۸۰۸٪
	SOS-C	۱۶.۲۵۶٪	۰.۰۱۵۷	۶۹.۷۸۴٪	۷۰.۴۸۰٪		۴۰.۹۲۷٪	۰.۰۱۳۲	۶۲.۹۵۶٪	۶۱.۱۴۷٪
	SSA-C	۱۶.۶۵۶٪	۰.۰۳۰۷	۷۰.۱۹۰٪	۷۱.۲۲۶٪		۴۲.۸۴۸٪	۰.۰۲۴۸	۶۰.۳۸۲٪	۵۸.۹۵۶٪
	۴GWO-C	۳.۶۱۹٪	۰.۰۰۹۶	۹۶.۵۴۷٪	۹۶.۴۶۴٪		۶.۳۰۱٪	۰.۰۱۹۶	۹۳.۲۴۷٪	۹۲.۵۸۱٪
	GWO-C	۷.۴۲۹٪	۰.۰۵۰۳	۹۲.۷۲۰٪	۹۲.۶۴۵٪		۵۲.۸۷۷٪	۰.۰۴۲۹	۴۴.۹۳۲٪	۴۷.۹۴۶٪
	PSO-C	۵.۷۱۴٪	۰.۰۲۴۴	۹۴.۴۰۵٪	۹۴.۳۴۵٪		۱۴.۲۴۷٪	۰.۰۵۲۱	۸۴.۷۶۱٪	۸۴.۸۸۱٪
	ABC-C	۹.۶۱۹٪	۰.۰۱۷۱	۹۰.۷۱۲٪	۹۰.۵۴۶٪		۵۲.۸۷۷٪	۰.۰۳۷۸	۴۶.۸۶۰٪	۴۸.۸۲۱٪
	SOS-C	۵.۳۸۱٪	۰.۰۱۱۲	۹۴.۷۴۷٪	۹۴.۶۸۳٪		۲۴.۲۸۴٪	۰.۰۵۴۳	۷۴.۵۵۶٪	۷۳.۳۷۵٪
	SSA-C	۶.۵۷۱٪	۰.۰۲۵۰	۹۳.۵۷۳٪	۹۳.۵۰۰٪		۳۱.۰۹۶٪	۰.۰۳۳۳	۶۷.۴۵۲٪	۶۷.۶۱۷٪
Balance scale	۴GWO-C	۱۱.۲۹۶٪	۰.۰۰۷۳	۸۹.۱۴۴٪	۸۸.۶۲۲٪	Tae	۹.۱۷۹٪	۰.۰۰۶۶	۹۴.۰۵۸٪	۸۷.۴۰۴٪
	GWO-C	۱۵.۷۷۸٪	۰.۰۱۹۸	۸۴.۷۹۸٪	۸۴.۱۲۲٪		۱۴.۳۵۹٪	۰.۰۲۱۴	۸۳.۴۶۹٪	۸۰.۱۱۸٪
	PSO-C	۱۲.۲۲۲٪	۰.۰۰۳۹	۸۸.۳۱۴٪	۸۷.۶۹۰٪		۹.۴۳۶٪	۰.۰۰۶۰	۹۳.۴۳۴٪	۸۶.۹۸۸٪
	ABC-C	۱۴.۳۳۳٪	۰.۰۰۵۰	۸۵.۹۵۶٪	۸۵.۵۴۰٪		۱۱.۱۲۸٪	۰.۰۰۹۱	۹۰.۹۶۱٪	۸۴.۴۱۲٪
	SOS-C	۱۱.۷۴۱٪	۰.۰۰۳۰	۸۸.۶۴۸٪	۸۸.۱۵۰٪		۹.۶۹۲٪	۰.۰۰۷۴	۹۳.۳۹۷٪	۸۶.۶۳۸٪
	SSA-C	۱۲.۷۴۱٪	۰.۰۰۹۸	۸۷.۳۸۴٪	۸۷.۱۲۰٪		۱۰.۶۶۷٪	۰.۰۱۳۰	۹۰.۳۲۸٪	۸۵.۰۰۶٪
	۴GWO-C	۰.۳۹۳٪	۰.۰۰۶۰	۹۹.۶۱۵٪	۹۹.۶۳۸٪		۱۳.۱۲۵٪	۰.۰۲۸۷	۶۶.۷۳۲٪	۶۶.۴۳۱٪
	GWO-C	۱۲.۴۷۲٪	۰.۰۵۸۴	۸۸.۱۸۳٪	۸۷.۷۹۷٪		۳۱.۵۷۷٪	۰.۰۳۱۶	۲۴.۳۶۰٪	۲۷.۵۱۹٪
	PSO-C	۱.۲۹۲٪	۰.۰۰۹۹	۹۸.۷۹۳٪	۹۸.۷۳۳٪		۱۸.۱۲۵٪	۰.۰۵۳۰	۵۵.۰۴۷٪	۵۶.۷۶۸٪
	ABC-C	۱۰.۲۲۵٪	۰.۰۱۶۳	۹۰.۶۰۵٪	۹۰.۲۷۲٪		۳۲.۱۱۳٪	۰.۰۱۵۱	۳۱.۳۶۶٪	۳۴.۵۱۹٪
	SOS-C	۲.۰۲۲٪	۰.۰۰۸۹	۹۷.۹۴۲٪	۹۸.۰۳۱٪		۲۱.۹۳۵٪	۰.۰۱۸۰	۴۶.۹۳۶٪	۴۷.۹۸۳٪
	SSA-C	۲.۴۱۶٪	۰.۰۱۵۹	۹۷.۶۵۴٪	۹۷.۶۴۳٪		۲۱.۰۴۲٪	۰.۰۴۵۵	۵۳.۰۸۷٪	۵۴.۰۰۵٪
Seeds	۴GWO-C	۳۴.۲۵۲٪	۰.۰۳۷۸	۴۱.۸۴۲٪	۴۳.۱۱۱٪	PenglungEW	۳۵.۶۰۳٪	۰.۰۳۶۶	۵۸.۳۸۳٪	۶۱.۱۶۹٪
	GWO-C	۴۶.۹۱۶٪	۰.۰۰۷۷	۲۱.۹۱۸٪	۲۲.۹۰۲٪		۵۷.۵۳۰٪	۰.۰۲۸۸	۲۶.۸۹۸٪	۳۲.۶۹۴٪
	PSO-C	۳۷.۶۶۴٪	۰.۰۴۱۳	۳۷.۵۷۲٪	۳۷.۷۶۷٪		۴۱.۴۶۶٪	۰.۰۳۲۲	۵۵.۴۹۹٪	۵۶.۹۱۴٪
	ABC-C	۴۴.۷۲۰٪	۰.۰۲۱۸	۲۷.۷۹۰٪	۲۸.۲۴۷٪		۵۳.۰۳۸٪	۰.۰۱۳۰	۳۸.۳۹۸٪	۴۲.۱۸۷٪
	SOS-C	۳۸.۸۷۹٪	۰.۰۳۴۵	۴۰.۸۹۱٪	۴۱.۳۰۹٪		۴۴.۳۹۷٪	۰.۰۲۲۵	۵۱.۸۲۹٪	۵۳.۷۶۴٪
	SSA-C	۴۱.۲۱۵٪	۰.۰۴۳۳	۳۵.۹۵۷٪	۳۵.۷۹۰٪		۴۴.۶۶۹٪	۰.۰۳۶۰	۵۴.۴۰۰٪	۵۴.۸۸۵٪

دادگان Zoo، Glass، و Tae بر اساس میانگین تابع برازندگی نمایش داده شده است. تابع برازندگی مورد استفاده در تحقیق از (۱۲) به دست می‌آید. این تابع باید کمینه گردد تا فاصله مراکز خوشه‌ها و داده‌های داخل آنها حداقل باشند. در بین الگوریتم‌ها، الگوریتم پیشنهادی کمترین مقدار و الگوریتم گرگ خاکستری بیشترین مقدار را به دست آورده است. در شکل ۷، نتایج الگوریتم‌های مورد استفاده برای خوشه‌بندی بر

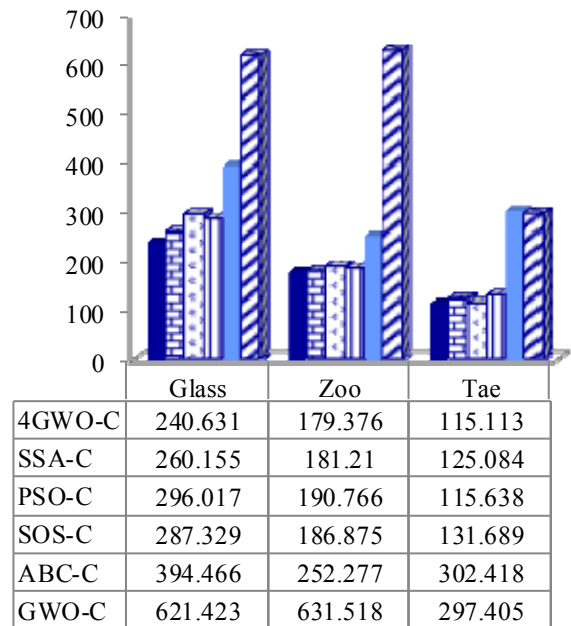
می‌رسیم، گرگ‌ها به بهینه نزدیک‌تر می‌گردند و بنابراین  $p_i$  آنها به یک نزدیک می‌شود و بر اساس حرکت سه گرگ برتر گروه، موقعیت بعدی آنها مشخص می‌گردد. در این حالت باز هم اگر گرگ‌ها در بهینه محلی باشند، چنانچه  $p_i$  گرگی کمتر از ۰/۵ باشد به اکتشاف می‌پردازد و امکان کشف فضای بهتر و فرار از بهینه محلی فراهم می‌گردد. در شکل ۸، نتایج الگوریتم‌ها برای خوشه‌بندی داده‌ها روی مجموعه

**Average F-measure%**



شکل ۷: نتایج الگوریتم‌های فراابتکاری بر اساس میانگین معیار  $F$  روی تمام مجموعه دادگان.

**Average Fitness Function**



شکل ۸: نتایج الگوریتم‌ها روی مجموعه دادگان Glass, Zoo, Tae بر اساس میانگین تابع برازندگی.

جدول ۶: نتایج حاصل از آزمون فریدمن بر اساس درصد خطای دسته‌بندی.

Dataset	Algorithm					
	4GWO-C	GWO-C	PSO-C	ABC-C	SOS-C	SSA-C
Iris	۲,۳	۲,۸۵	۳,۹	۵,۵۵	۲,۶۵	۳,۷۵
Breast cancer Wisconsin (Original)	۲,۳	۳,۸	۳,۴	۴,۸۵	۲,۰۵	۴,۶
Balance scale	۱,۷	۳,۱	۴,۸	۵,۸	۲,۷	۲,۹
Seeds	۱,۷	۳,۵۵	۳,۲۵	۵,۳۵	۳,۳۵	۳,۸
Statlog (Heart)	۱,۶	۵,۶	۳,۰۵	۵,۳	۲,۰۵	۳,۴
Wine	۱,۳	۵,۵۵	۲,۳۵	۵,۴۵	۳,۲	۳,۱۵
Glass	۱,۳	۵,۶۵	۲,۶	۴,۸۵	۲,۹	۳,۷
Zoo	۱,۰۵	۵,۷	۲,۹۵	۵,۳	۲,۸	۳,۲
Solar	۱	۶	۳,۰۵	۴,۸	۲,۳	۳,۸۵
Tae	۲,۴	۴,۲۵	۲,۴	۳,۸۵	۳,۴	۴,۷
PenglungEW	۱,۰۵	۵,۶	۲,۱۵	۵,۴	۲,۹۵	۳,۸۵
Parkinson Disease	۱,۸	۵,۷۵	۲,۲۵	۴,۵۵	۲,۸	۳,۸۵
Ecoli	۱,۳۵	۵,۴	۲,۰۵	۵,۵	۳,۶	۳,۱
Vehicle	۱,۲	۵,۸	۲,۱۵	۵,۲	۳,۱۵	۳,۵
Sum	۲۲,۰۵	۶۸,۶	۴۰,۳۵	۷۱,۷۵	۳۹,۹	۵۱,۳۵
Rank	۱	۵	۳	۶	۲	۴

است [۵۱]. از روی رتبه‌بندی این جدول، الگوریتم پیشنهادی بهترین نتایج را نشان می‌دهد و در رده‌های بعدی، الگوریتم ارگانیسم‌های هم‌زیست و ازدحام ذرات قرار دارند.

در جدول ۷، کارایی الگوریتم پیشنهادی 4GWO-C روی مسأله خوشه‌بندی با عملکرد الگوریتم گرگ خاکستری و چند نسخه بهبودیافته الگوریتم بهینه‌سازی گرگ خاکستری نظیر AGWO [۴۵]، EEGWO [۴۶]، EGWO [۴۷] و G-NHGWO [۴۸] مقایسه گردیده است. این الگوریتم‌ها برای مسأله خوشه‌بندی به ترتیب AGWO-C [۴۵]، EEGWO-C [۴۶]، EGWO-C [۴۷] و G-NHGWO-C نامیده

اساس میانگین معیار  $F$ ، روی تمام مجموعه دادگان مقایسه گردیده‌اند. در این شکل الگوریتم پیشنهادی بهترین نتیجه را به دست آورده و با اختلاف زیادی، الگوریتم بهینه‌سازی ازدحام ذرات در رتبه دوم است. مطابق این شکل، الگوریتم گرگ خاکستری عملکرد ضعیف‌تری نسبت به همه الگوریتم‌ها بر اساس معیار  $F$  داشته و اختلاف آن با الگوریتم پیشنهادی بسیار قابل توجه (۱۶/۳۰۴٪) است.

جدول ۸، نتایج حاصل از آزمون فریدمن را بر اساس درصد خطای دسته‌بندی نشان می‌دهد. مقادیر کمتر در این جدول بهتر هستند. آزمون فریدمن، آزمونی ناپارامتری آماری برای مقایسه میانگین رتبه‌ها در گروه‌ها

جدول ۷: نتایج حاصل از اجرای الگوریتم روی مجموعه دادگان.

Metric	Dataset	MCR	STD	Precision	F-Measure	Dataset	MCR	STD	Precision	F-Measure
Algorithm										
Iris	FGWO-C	۰.۹۳۳٪	۰.۰۰۳۴	۹۹.۰۸۰٪	۹۹.۰۷۳٪	Zoo	۱.۸۸۱٪	۰.۰۱۵۸	۹۶.۱۳۷٪	۹۵.۴۴۰٪
	GWO-C	۱.۱۳۳٪	۰.۰۰۳۲	۹۸.۸۸۶٪	۹۸.۸۷۶٪		۲۱.۶۸۳٪	۰.۰۳۴۸	۴۸.۰۰۵٪	۵۱.۱۷۱٪
	G-NHGWOC	۱.۰۰۰٪	۰.۰۰۳۵	۹۹.۰۱۷٪	۹۹.۰۰۸٪		۲۳.۴۶۵٪	۰.۰۴۱۲	۴۰.۷۹۷٪	۴۴.۸۰۹٪
	AGWO-C	۱.۴۰۰٪	۰.۰۰۵۸	۹۸.۶۴۹٪	۹۸.۶۲۵٪		۲۴.۳۵۶٪	۰.۰۳۱۷	۴۵.۳۳۵٪	۴۷.۷۷۰٪
	EGWO-C	۱.۲۶۷٪	۰.۰۰۶۶	۹۸.۷۸۰٪	۹۸.۷۵۷٪		۴.۰۵۹٪	۰.۰۲۸۲	۹۱.۹۴۴٪	۹۱.۰۰۶٪
	EEGWO-C	۴.۵۳۳٪	۰.۰۱۶۶	۹۵.۶۱۳٪	۹۵.۵۴۰٪		۲۴.۶۵۴٪	۰.۰۲۲۶	۴۵.۰۲۷٪	۴۸.۳۷۶٪
	FGWO-C	۱.۹۴۷٪	۰.۰۰۱۲	۹۷.۵۴۵٪	۹۷.۸۹۲٪		۱۱.۰۱۰٪	۰.۰۱۱۹	۸۹.۲۸۵٪	۸۹.۰۵۸٪
	GWO-C	۲.۱۲۳٪	۰.۰۰۲۱	۹۷.۲۵۳٪	۹۷.۷۲۱٪		۲۸.۲۲۱٪	۰.۰۲۴۹	۷۲.۲۰۶٪	۷۱.۸۷۲٪
	G-NHGWOC	۲.۱۸۲٪	۰.۰۰۱۸	۹۷.۱۷۷٪	۹۷.۶۶۰٪		۲۷.۲۱۲٪	۰.۰۲۷۵	۷۳.۶۹۰٪	۷۳.۰۹۳٪
	AGWO-C	۲.۱۶۷٪	۰.۰۰۲۳	۹۷.۲۵۴٪	۹۷.۶۶۲٪		۲۷.۷۸۹٪	۰.۰۲۰۰	۷۲.۶۲۱٪	۷۲.۲۸۱٪
Breast Cancer Wisconsin	EGWO-C	۲.۰۵۰٪	۰.۰۰۱۴	۹۷.۴۱۰٪	۹۷.۷۸۵٪	Solar	۱۳.۴۶۲٪	۰.۰۲۶۳	۸۶.۹۲۳٪	۸۶.۶۷۵٪
	EEGWO-C	۲.۹۸۷٪	۰.۰۰۳۵	۹۶.۴۶۹٪	۹۶.۷۵۲٪		۲۸.۲۲۱٪	۰.۰۲۳۶	۷۳.۴۱۹٪	۷۲.۵۲۸٪
	FGWO-C	۱۴.۷۶۸٪	۰.۰۱۴۴	۷۰.۸۰۴٪	۷۱.۲۵۰٪		۳۹.۲۷۲٪	۰.۰۲۱۴	۶۳.۴۷۶٪	۶۲.۲۶۹٪
	GWO-C	۱۶.۶۵۶٪	۰.۰۱۷۱	۶۹.۲۱۶٪	۶۹.۹۱۰٪		۴۳.۵۱۰٪	۰.۰۵۰۲	۵۷.۳۶۷٪	۵۶.۸۶۴٪
	G-NHGWOC	۱۵.۶۸۰٪	۰.۰۲۳۳	۷۰.۲۴۱٪	۷۱.۱۰۵٪		۴۲.۰۵۳٪	۰.۰۳۸۴	۶۱.۳۱۰٪	۵۹.۷۴۷٪
	AGWO-C	۱۷.۹۵۲٪	۰.۰۱۵۰	۶۸.۴۵۷٪	۶۹.۲۸۵٪		۴۳.۵۷۶٪	۰.۰۳۸۱	۵۶.۷۸۳٪	۵۶.۵۷۷٪
	EGWO-C	۱۵.۴۰۸٪	۰.۰۱۷۷	۷۰.۶۲۵٪	۷۱.۵۰۵٪		۴۱.۳۹۱٪	۰.۰۱۸۵	۶۳.۰۰۳٪	۶۰.۹۷۰٪
	EEGWO-C	۳۵.۸۰۸٪	۰.۰۳۶۸	۶۳.۵۲۶٪	۶۲.۲۷۳٪		۴۶.۶۲۳٪	۰.۰۱۴۴	۵۷.۲۱۵٪	۵۵.۴۴۰٪
	FGWO-C	۳.۶۱۹٪	۰.۰۰۹۶	۹۶.۵۴۷٪	۹۶.۴۶۴٪		۶.۳۰۱٪	۰.۰۱۹۶	۹۳.۲۴۷٪	۹۲.۵۸۱٪
	GWO-C	۷.۴۲۹٪	۰.۰۵۰۳	۹۲.۷۲۰٪	۹۲.۶۴۵٪		۵۲.۸۷۷٪	۰.۰۴۲۹	۴۴.۹۳۲٪	۴۷.۹۴۶٪
Balance scale	G-NHGWOC	۵.۴۸۱٪	۰.۰۲۴۶	۹۴.۷۵۳٪	۹۴.۶۸۶٪	PenglungEW	۵۵.۶۱۶٪	۰.۰۵۱۳	۴۱.۸۹۵٪	۴۴.۳۵۷٪
	AGWO-C	۱۲.۴۲۹٪	۰.۰۵۹۲	۸۸.۳۳۳٪	۸۷.۹۴۸٪		۵۶.۱۶۴٪	۰.۰۲۹۶	۴۰.۲۹۹٪	۴۴.۸۰۲٪
	EGWO-C	۵.۱۹۰٪	۰.۰۳۰۸	۹۴.۹۲۳٪	۹۴.۸۶۶٪		۱۱.۲۳۳٪	۰.۰۳۷۰	۸۸.۳۰۲٪	۸۷.۸۵۵٪
	EEGWO-C	۲۴.۵۳۴٪	۰.۰۹۲۸	۷۷.۹۸۳٪	۷۶.۴۸۲٪		۵۸.۲۱۹٪	۰.۰۴۵۳	۴۴.۷۱۴٪	۴۵.۱۲۱٪
	FGWO-C	۱۱.۲۹۶٪	۰.۰۰۷۳	۸۹.۱۴۴٪	۸۸.۶۲۲٪		۹.۱۷۹٪	۰.۰۰۶۶	۹۴.۰۵۸٪	۸۷.۴۰۴٪
	GWO-C	۱۵.۷۷۸٪	۰.۰۱۹۸	۸۴.۷۹۸٪	۸۴.۱۲۲٪		۱۴.۳۵۹٪	۰.۰۲۱۴	۸۳.۴۹۶٪	۸۰.۱۱۸٪
	G-NHGWOC	۱۶.۳۷۰٪	۰.۰۲۲۵	۸۳.۸۱۲٪	۸۲.۴۵۳٪		۱۴.۳۰۸٪	۰.۰۲۳۹	۸۳.۹۵۴٪	۸۰.۶۷۷٪
	AGWO-C	۱۷.۸۵۲٪	۰.۰۲۷۰	۸۲.۳۸۸٪	۸۲.۱۱۸٪		۱۴.۸۲۱٪	۰.۰۲۲۶	۸۱.۸۹۳٪	۷۹.۷۸۹٪
	EGWO-C	۱۱.۵۱۹٪	۰.۰۰۴۱	۸۸.۹۰۲٪	۸۸.۳۸۱٪		۹.۷۴۴٪	۰.۰۱۰۰	۹۳.۵۸۹٪	۸۶.۵۷۸٪
	EEGWO-C	۱۸.۴۸۳٪	۰.۰۱۲۶	۸۱.۴۸۰٪	۸۱.۲۹۷٪		۱۲.۵۱۳٪	۰.۰۱۱۱	۸۷.۵۹۰٪	۸۲.۲۵۳٪
Statlog (Heart)	FGWO-C	۰.۳۹۳٪	۰.۰۰۶۰	۹۹.۶۱۵٪	۹۹.۶۳۸٪	Parkinson	۱۳.۱۲۵٪	۰.۰۲۸۷	۶۶.۷۳۲٪	۶۶.۴۳۱٪
	GWO-C	۱۲.۴۷۲٪	۰.۰۵۸۴	۸۸.۱۸۳٪	۸۷.۷۹۷٪		۳۱.۵۷۷٪	۰.۰۳۱۶	۲۴.۳۶۰٪	۲۷.۵۱۹٪
	G-NHGWOC	۱۵.۶۱۸٪	۰.۰۷۳۵	۸۳.۰۷۰٪	۸۳.۴۹۷٪		۲۹.۳۷۵٪	۰.۰۴۰۸	۲۸.۴۷۷٪	۳۲.۳۲۴٪
	AGWO-C	۱۲.۰۲۳٪	۰.۰۷۳۹	۸۸.۹۴۱٪	۸۸.۵۸۹٪		۳۱.۰۴۲٪	۰.۰۴۲۱	۲۷.۸۹۴٪	۳۱.۰۸۴٪
	EGWO-C	۰.۴۴۹٪	۰.۰۰۵۸	۹۹.۵۴۲٪	۹۹.۵۷۲٪		۱۵.۶۵۵٪	۰.۰۳۵۲	۶۹.۱۴۲٪	۶۶.۹۸۴٪
	EEGWO-C	۱۸.۸۷۶٪	۰.۰۵۵۵	۸۴.۵۰۴٪	۸۳.۰۳۱٪		۳۳.۵۱۲٪	۰.۰۰۸۱	۲۷.۰۲۰٪	۲۹.۶۴۰٪
	FGWO-C	۳۴.۲۵۲٪	۰.۰۳۷۸	۴۱.۸۴۴٪	۴۳.۱۱۱٪		۳۵.۶۰۳٪	۰.۰۳۶۶	۵۸.۳۸۳٪	۶۱.۱۶۹٪
	GWO-C	۴۶.۹۱۶٪	۰.۰۰۷۷	۲۱.۹۱۸٪	۲۲.۹۰۲٪		۵۷.۵۳۰٪	۰.۰۲۸۸	۲۶.۸۹۸٪	۳۲.۶۹۴٪
	G-NHGWOC	۴۶.۱۶۸٪	۰.۰۲۸۱	۲۰.۱۵۶٪	۲۲.۵۶۶٪		۵۲.۲۳۴٪	۰.۰۶۱۵	۲۹.۶۱۲٪	۳۶.۴۸۷٪
	AGWO-C	۴۷.۲۴۳٪	۰.۰۱۰۹	۱۹.۰۱۰٪	۲۰.۴۲۶٪		۵۴.۶۳۴٪	۰.۰۲۴۰	۲۶.۹۸۰٪	۳۳.۸۵۸٪
Wine	EGWO-C	۳۷.۰۰۹٪	۰.۰۴۴۸	۲۸.۷۵۵٪	۲۹.۵۴۱٪	Vehicle	۳۷.۸۱۳٪	۰.۰۲۸۱	۵۵.۵۶۳٪	۵۸.۷۰۲٪
	EEGWO-C	۴۷.۲۹۰٪	۰.۰۱۲۸	۲۳.۰۱۰٪	۲۴.۵۱۵٪		۵۵.۹۱۰٪	۰.۰۲۸۸	۳۵.۹۲۷٪	۳۹.۴۸۷٪

می‌شوند و روی مجموعه دادگان جدول ۳ با یکدیگر به رقابت می‌پردازند. همان گونه که در جدول ۷ مشخص است الگوریتم پیشنهادی، نتایج بهتری را در این جدول در مقایسه با سایر الگوریتم‌ها نشان می‌دهد. روش پیشنهادی در مجموعه دادگانی چون Zoo و PenglungEW با اختلاف زیادی، عملکرد بهتری از نظر معیارهای مورد مقایسه دارد. در این مجموعه دادگان روش پیشنهادی به ترتیب دقتی برابر با ۹۶/۱۳۷٪ و

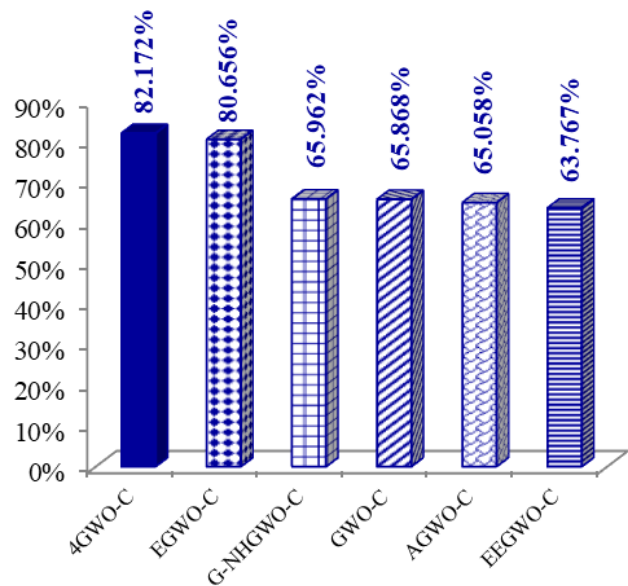
۹۳/۲۴۷٪ نشان می‌دهد، در صورتی که الگوریتم EGWO-C که در رتبه بعدی قرار دارد، به ترتیب دقتی برابر با ۹۱/۹۹۴٪ و ۸۸/۳۰۲٪ روی این مجموعه دادگان نشان می‌دهد. تعداد مراکز خوشه در این مجموعه دادگان ۷ می‌باشد که غیر از روش پیشنهادی و EGWO-C، سایر الگوریتم‌های بهبودیافته گرگ نتوانسته‌اند نتایج قابل قبولی را ارائه دهند و دقتی زیر ۵۰٪ داشته‌اند. در بین نسخه‌های مورد مقایسه، الگوریتم EEGWO-C

مجموعه دادگان نامتوازن، استفاده از الگوریتم پیشنهادی در هسته بسیاری از روش‌هایی که نیاز به خوشه‌بندی داده‌ها دارند و یا الگوریتم‌هایی مانند الگوریتم فاخته که در هسته خود از خوشه‌بندی استفاده می‌کنند، استفاده از روابط روش پیشنهادی برای بهبود الگوریتم گرگ خاکستری در سایر الگوریتم‌های فراابتکاری و ارزیابی عملکرد آنها و پیاده‌سازی الگوریتم دودویی و چندهدفه از الگوریتم ۴GWO.

## مراجع

- [1] M. Jafarzagdegan, F. Safi-Esfahani, and Z. Beheshti, "Combining hierarchical clustering approaches using the PCA method," *Expert Syst. Appl.*, vol. 137, pp. 1-10, Dec 2019.
- [2] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
- [3] D. J. Hand, "Principles of data mining," *Drug Saf.*, vol. 30, no. 7, pp. 621-622, Jul. 2007.
- [4] D. T. Dinh and V. N. Huynh, "k-PbC: an improved cluster center initialization for categorical data clustering," *Appl. Intell.*, vol. 50, no. 8, pp. 2610-2632, Aug. 2020.
- [5] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645-678, May 2005.
- [6] Z. Beheshti, "A novel x-shaped binary particle swarm optimization," *Soft Comput.*, vol. 25, no. 4, pp. 3013-3042, Feb. 2021.
- [7] Z. Beheshti, "UTF: upgrade transfer function for binary meta-heuristic algorithms," *Appl. Soft Comput.*, vol. 106, Article ID: 107346, 28 pp., Jul. 2021.
- [8] R. Salgotra, U. Singh, S. Singh, G. Singh, and N. Mittal, "Self-adaptive salp swarm algorithm for engineering optimization problems," *Appl. Math. Model.*, vol. 89, pp. 188-207, Jul. 2021.
- [9] Z. Beheshti, "BMNABC: binary multi-neighborhood artificial bee colony for high-dimensional discrete optimization problems," *Cybern. Syst.*, vol. 49, no. 7-8, pp. 452-474, Nov. 2018.
- [10] Z. Beheshti, S. M. Shamsuddin, S. Hasan, and N. E. Wong, "Improved centripetal accelerated particle swarm optimization," *Int. J. Adv. Soft Comput. its Appl.*, vol. 8, no. 2, pp. 1-26, Jul. 2016.
- [11] I. Aljarah, M. Mafarja, A. A. Heidari, H. Faris, and S. Mirjalili, "Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach," *Knowl. Inf. Syst.*, vol. 62, no. 2, pp. 507-539, Feb. 2020.
- [12] H. D. Menendez, F. E. B. Otero, and D. Camacho, "Medoid-based clustering using ant colony optimization," *Swarm Intell.*, vol. 10, no. 2, pp. 123-145, Jul. 2016.
- [13] P. Das, D. K. Das, and S. Dey, "A modified bee colony optimization (MBCO) and its hybridization with k-means for an application to data clustering," *Appl. Soft Comput.*, vol. 70, pp. 590-603, Sept. 2018.
- [14] R. J. Kuo and F. E. Zulvia, "An improved differential evolution with cluster decomposition algorithm for automatic clustering," *Soft Comput.*, vol. 23, no. 18, pp. 8957-8973, Sept. 2019.
- [15] T. Ashish, S. Kapi, and B. Manju, "Parallel bat algorithm-based clustering using mapreduce," In: G. Perez, K. Mishra, S. Tiwari, and M. Trivedi (eds), *Networking Communication and Data Knowledge Engineering, Lecture Notes on Data Engineering and Communications Technologies*, vol 4. Springer, pp. 73-82, 2018.
- [16] O. Tarkhaneh, A. Isazadeh, and H. J. Khamnei, "A new hybrid strategy for data clustering using cuckoo search based on Mantegna levy distribution, PSO and k-means," *Int. J. Comput. Appl. Technol.*, vol. 58, no. 2, pp. 137-149, Jan. 2018.
- [17] H. A. Abdulwahab, A. Noraziah, A. A. Alsewari, and S. Q. Salih, "An enhanced version of black hole algorithm via levy flight for optimization and data clustering problems," *IEEE Access*, vol. 7, pp. 142085-142096, Aug. 2019.
- [18] K. W. Huang, Z. X. Wu, H. W. Peng, M. C. Tsai, Y. C. Hung, and Y. C. Lu, "Memetic particle gravitation optimization algorithm for solving clustering problems," *IEEE Access*, vol. 7, pp. 80950-80968, Jan. 2019.
- [19] H. Yu, Z. Chang, G. Wang, and X. Chen, "An efficient three-way clustering algorithm based on gravitational search," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 1003-1016, May 2020.
- [20] A. Kaur, S. K. Pal, and A. P. Singh, "Hybridization of chaos and flower pollination algorithm over K-means for data clustering," *Appl. Soft Comput.*, vol. 97, Article ID: 105523, 17 pp., Dec. 2019.
- [21] Y. Zhou, H. Wu, Q. Luo, and M. Abdel-Baset, "Automatic data clustering using nature-inspired symbiotic organism search

## Average F-measure%



شکل ۸: نتایج الگوریتم‌های بهبودیافته گرگ خاکستری بر اساس میانگین معیار  $F$  روی تمام مجموعه دادگان.

ضعیف‌ترین عملکرد را دارد. این عملکرد ضعیف در خطای به دست آمده روی مجموعه دادگان Seeds و Balance scale بسیار قابل توجه است.

شکل ۸، نتایج الگوریتم‌های بهبودیافته گرگ خاکستری را روی تمام مجموعه دادگان بر اساس میانگین معیار  $F$  نشان می‌دهد. در این شکل، الگوریتم پیشنهادی چهارگرگ بهترین نتیجه را به دست آورده و الگوریتم EGWO در رتبه دوم قرار دارد. بر اساس این شکل، الگوریتم EEGWO بدترین عملکرد را نسبت به سایر الگوریتم‌های بهبودیافته گرگ خاکستری در معیار  $F$  دارد. در کل، سایر نسخه‌های الگوریتم گرگ خاکستری یعنی AGWO و G-NHGWOC نیز کارایی خوبی ندارند.

## ۵- نتیجه‌گیری و تحقیقات آینده

در این تحقیق، ابتدا به منظور حل مشکل همگرایی زودرس و گرفتارشدن در بهینه محلی توسط الگوریتم بهینه‌سازی گرگ خاکستری، الگوریتم بهبودیافته‌ای ارائه شد. الگوریتم پیشنهادی بر روی مسأله خوشه‌بندی پیاده‌سازی گردید و نتایج خوشه‌بندی با الگوریتم پیشنهادی با الگوریتم‌هایی از جمله الگوریتم گرگ خاکستری، الگوریتم زنبور عسل مصنوعی، الگوریتم ارگانیسم‌های همزیست، الگوریتم سالپ و چند نسخه بهبودیافته از الگوریتم گرگ خاکستری بر اساس درصد طبقه‌بندی نادرست (خطا)، انحراف معیار، دقت و معیار  $F$  با یکدیگر مقایسه شده‌اند. در مسأله خوشه‌بندی، مجموعه دادگان چالشی با تعداد مراکز خوشه زیاد و داده‌های نامتوازن استفاده شد. نتایج جداول، نمودارها و همچنین آزمون فریدمن نشان‌دهنده عملکرد بسیار خوب الگوریتم پیشنهادی می‌باشد. الگوریتم پیشنهادی، قابلیت اکتشاف و توانایی فرار از بهینه‌های محلی را در الگوریتم گرگ خاکستری بهبود داده و همچنین عملکردی بهتر از الگوریتم مورد مقایسه نشان می‌دهد.

هرچند الگوریتم پیشنهادی دارای عملکرد بالایی در مسأله خوشه‌بندی است، می‌توان در جهت گسترش این تحقیق در خصوص راهکارهای آینده به مواردی اشاره کرد از جمله: ارزیابی عملکرد الگوریتم پیشنهادی ۴GWO در سایر مسائلی بهینه‌سازی، استفاده از روش پیشنهادی در خوشه‌بندی و ترکیب آن با طبقه‌بندی برای آموزش نیمه‌نظارتی در

- [۳۹] م. یقینی و ج. لسان، "حل مسأله خوشه‌بندی ظرفیت‌دار با استفاده از روش‌های مبتنی بر الگوریتم‌های شبیه‌سازی تیریدی و ژنتیک،" *نشریه بین‌المللی مهندسی صنایع و مدیریت تولید*، دوره ۲۱، شماره ۳، صص. ۵۴-۴۵، پاییز ۱۳۸۹.
- [۴۰] ع. فهمی جعفرقلخانلو و م. شمسی، "بخش‌بندی تصاویر رنگی چهره مبتنی بر خوشه‌بند فازی بهینه‌سازی شده با الگوریتم‌های گرگ خاکستری و نهنگ،" *مجله علمی رایانش نرم و فناوری اطلاعات*، دوره ۱۰، شماره ۲، صص. ۱۳-۱۰، تیر ۱۴۰۰.
- [41] Y. Li, X. Lin, and J. Liu, "An improved gray wolf optimization algorithm to solve engineering problems," *Sustainability*, vol. 13, no. 6, Article ID: 3208, 23 pp., 15 Mar. 2021.
- [42] M. Kohli and S. Arora, "Chaotic grey wolf optimization algorithm for constrained optimization problems," *J. Comput. Des. Eng.*, vol. 5, no. 4, pp. 458-472, Oct. 2018.
- [43] Z. Yue, S. Zhang, and W. Xiao, "A novel hybrid algorithm based on grey wolf optimizer and fireworks algorithm," *Sensors*, vol. 20, no. 7, Article ID: 2147, 17 pp., Jan. 2020.
- [۴۴] ع. محمدزاده، م. مصدری، ف. سلیمانیان قرچقیق و ا. جعفریان، "ارائه یک الگوریتم بهبودیافته بهینه‌سازی گرگ‌های خاکستری برای زمان‌بندی جریان کار در محیط محاسبات ابری،" *مجله علمی رایانش نرم و فناوری اطلاعات*، دوره ۸، شماره ۴، صص. ۲۹-۱۷، بهمن ۱۳۹۸.
- [45] W. Long, J. Jiao, X. Liang, and M. Tang, "An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization," *Eng. Appl. Artif. Intell.*, vol. 68, pp. 63-80, Feb. 2018.
- [46] M. H. Qais, H. M. Hasanien, and S. Alghuwainem, "Augmented grey wolf optimizer for grid-connected PMSG-based wind energy conversion systems," *Appl. Soft Comput.*, vol. 69, pp. 504-515, Aug. 2018.
- [47] K. Luo, "Enhanced grey wolf optimizer with a model for dynamically estimating the location of the prey," *Appl. Soft Comput.*, vol. 77, no. ???, pp. 225-235, Apr. 2019.
- [48] E. Akbari, A. Rahimnejad, and S. A. Gadsden, "A greedy non-hierarchical grey wolf optimizer for real-world optimization," *Electron. Lett.*, vol. 57, no. 13, pp. 499-501, Jan. 2021.
- [49] A. Bahrololoum, H. Nezamabadipour, and S. Saryazdi, "A data clustering approach based on universal gravity rule," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 415-428, Oct. 2015.
- [50] D. Dua and C. Graff, {UCI} Machine Learning Repository, 2017.
- [51] J. Derrac, S. Garcia, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 3-18, Mar. 2011.
- لااله عجمی بختیاروند تحصیلات کارشناسی و کارشناسی ارشد خود را در رشته مهندسی کامپیوتر گرایش نرم افزار در دانشگاه آزاد اسلامی واحد نجف آباد به پایان رساند. زمینه‌های پژوهشی مورد علاقه وی، محاسبات نرم، الگوریتم‌های فراابتکاری و داده کاوی می‌باشد.
- زهرا بهشتی تحصیلات دکتری و پسا دکتری خود را در رشته علوم کامپیوتر گرایش هوش مصنوعی در دانشگاه تکنولوژی مالزی به پایان رساند و هم‌اکنون عضو هیات علمی دانشکده مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد نجف آباد می‌باشد. وی بر اساس اطلاعات و آمار محققان دانشگاه استنفورد، با در نظر گرفتن شاخص‌های استنادی استاندارد، در فهرست دانشمندان دو درصد برتر جهان، در سال ۲۰۲۰ قرار گرفت. زمینه‌های پژوهشی مورد علاقه ایشان، الگوریتم‌های فراابتکاری، یادگیری ماشین، داده کاوی و تحلیل کلان داده می‌باشد.
- algorithm," *Knowledge-Based Syst.*, vol. 163, pp. 546-557, Jan. 2019.
- [22] G. Drakopoulos, et al., "A genetic algorithm for spatio-social tensor clustering," *Evol. Syst.*, vol. 11, no. 3, pp. 1-11, Sept. 2019.
- [23] M. Alswaiti, M. Albughdadi, and N. A. M. Isa, "Density-based particle swarm optimization algorithm for data clustering," *Expert Syst. Appl.*, vol. 91, pp. 170-186, Jan. 2018.
- [24] J. Chen, X. Qi, L. Chen, F. Chen, and G. Cheng, "Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection," *Knowledge-Based Syst.*, vol. 203, Article ID: 106167, 10 pp., Sept. 2020.
- [25] L. M. Abualigah, A. T. Khader, E. S. Hanandeh, and A. H. Gandomi, "A novel hybridization strategy for krill herd algorithm applied to clustering techniques," *Appl. Soft Comput.*, vol. 60, pp. 423-435, Nov. 2017.
- [26] M. Demri, S. Ferouhat, S. Zakaria, and M. E. Barmati, "A hybrid approach for optimal clustering in wireless sensor networks using cuckoo search and simulated annealing algorithms," in *Proc. 2nd Int. Conf. on Mathematics and Information Technology, ICMIT'20*, pp. 202-207, Adrar, Algeria, 18-19 Feb. 2020.
- [27] S. Gavel, P. Joshi, S. Tiwari, and A. S. Raghuvanshi, "An optimized hybrid clustering method using salp swarm optimization with K-means," In: V. Nath and J. Mandal (eds) *Nanoelectronics, Circuits and Communication Systems, Lecture Notes in Electrical Engineering*, vol. 692, Springer, pp. 247-254, 2020.
- [28] S. K. Majhi, "Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection," *Evol. Intell.*, vol. 14, no. 1, pp. 35-46, 2021.
- [29] E. Rashedi, H. Nezamabadipour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Inf. Sci. (Nij)*, vol. 179, no. 13, pp. 2232-2248, Jan. 2009.
- [30] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46-61, 2014.
- [31] G. P. Gupta and S. Jha, "Integrated clustering and routing protocol for wireless sensor networks using Cuckoo and Harmony Search based metaheuristic techniques," *Eng. Appl. Artif. Intell.*, vol. 68, pp. 101-109, Feb. 2018.
- [32] G. P. Gupta and B. Saha, "Load balanced clustering scheme using hybrid metaheuristic technique for mobile sink based wireless sensor networks," *J. Ambient Intell. Humaniz. Comput.*, 12 pp., Apr. 2020. (in Press)
- [33] M. Shakil, A. Fuad Yousif Mohammed, R. Arul, A. K. Bashir, and J. K. Choi, "A novel dynamic framework to detect DDoS in SDN using metaheuristic clustering," *Trans. Emerg. Telecommun. Technol.*, 12 pp., Apr. 2019. (in Press)
- [34] R. J. Kuo, T. C. Lin, F. E. Zulvia, and C. Y. Tsai, "A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis," *Appl. Soft Comput.*, vol. 67, pp. 299-308, Jan. 2018.
- [35] A. Kaur and Y. Kumar, "A new metaheuristic algorithm based on water wave optimization for data clustering," *Evol. Intell.*, 12 pp., Jan. 2021.
- [36] M. Ghobaei-Arani and A. Shahidinejad, "An efficient resource provisioning approach for analyzing cloud workloads: a metaheuristic-based clustering approach," *J. Supercomput.*, vol. 77, no. 1, pp. 711-750, 2021.
- [۳۷] ج. املشی و ی. بستانی املشی، "روشی جدید به منظور خوشه‌بندی داده‌های سرعت باد در نیروگاه‌های بادی با استفاده از الگوریتم‌های FCM و PSO،" *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، سال ۸، شماره ۳، صص. ۲۱۴-۲۱۰، پاییز ۱۳۸۹.
- [۳۸] م. ع. نژاد و م. خرد، "استفاده از خوشه‌بندی BIRCH و الگوریتم بهینه‌سازی واکنش شیمیایی جهت کشف تقلب در حوزه سلامت،" *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، ب- مهندسی کامپیوتر، سال ۱۷، شماره ۲، صص. ۱۶۰-۱۵۳، تابستان ۱۳۹۸.