

# انتخاب ویژگی چندبرچسب با استفاده از راهکار ترکیبی مبتنی بر الگوریتم بهینه‌سازی ازدحام ذرات

آذر رفیعی، پرهام مرادی و عبدالباقی قادرزاده

مربوط به انتشار یک دستگاه آیفون جدید با برچسب شرکت و برچسب دیگر مربوط به تکنولوژی حاشیه‌نویسی می‌باشد و به عبارت دیگر، هر نمونه به جای یک برچسب با مجموعه‌ای از برچسب‌ها در ارتباط است. یادگیری چندبرچسبی یکی از حوزه‌های یادگیری ماشین است که به یادگیری داده‌های چندبرچسبی می‌پردازد و در آن، هر نمونه با برچسب‌های چندگانه بالقوه ارتباط دارد. روش‌های موجود طبقه‌بندی چندبرچسبی را می‌توان به دو دسته اصلی تقسیم‌بندی کرد: در روش اول، مسأله یادگیری چندبرچسبی به یک یا چند مسأله تک‌برچسبی تبدیل می‌گردد تا بتوان از الگوریتم‌های یادگیری تک‌برچسبی استفاده نمود. در دومین رویکرد، الگوریتم‌های یادگیری به طور مستقیم با داده‌های چندبرچسبی کار می‌کنند. یکی از مشکلات راهکارهای طبقه‌بندی چندبرچسبی این است که کارایی آنها با افزایش ابعاد مسأله و ورودی‌ها به شدت کاهش پیدا می‌کند. انتخاب ویژگی، یکی از راهکارهای اصلی برای کاهش ابعاد مسأله در این نوع مسائل است. هدف از روش‌های انتخاب ویژگی، کاهش ابعاد مسأله با انتخاب ویژگی‌های مرتبط و غیر افزونه و در نظر نگرفتن سایر ویژگی‌ها است. انتخاب ویژگی منجر به کاهش ورودی‌ها و افزایش کارایی الگوریتم‌های طبقه‌بندی چندبرچسبی می‌شود. داده‌های واقعی ممکن است که حاوی ویژگی‌هایی با تأثیرات متفاوت در تعیین برچسب کلاس مسأله باشند. برای مثال، جنسیت یک فرد برای پیش‌بینی یک برچسب بیماری مانند "دیابت" در مقایسه با سن او، ارتباط کمتری دارد. ویژگی‌های نامربوط، معمولاً به صحت مدل طبقه‌بندی به‌علاوه ناکارآمدی محاسبات خدشه وارد می‌کنند و در نتیجه، ایده اصلی انتخاب ویژگی، انتخاب زیرمجموعه‌ای از ویژگی‌های مطلوب با از بین بردن ویژگی‌های نامربوط با افزونگی بالا است. این کاهش به تسریع فرایند یادگیری کمک می‌کند. به عبارت دیگر، انتخاب ویژگی یک راهکار برای حل مسأله ازدحام ابعاد است که طی آن ویژگی‌های اضافی و نامربوط از مجموعه ویژگی‌های موجود حذف می‌شوند [۱].

تا کنون برای انتخاب ویژگی، روش‌های متنوعی استفاده شده که به طور کلی می‌توان آنها را به سه دسته اصلی تقسیم‌بندی کرد: روش‌های فیلتری [۲] تا [۷]، روش‌های پوششی [۸] تا [۱۲] و روش‌های ترکیبی [۱۳] و [۱۴]. روش‌های فیلتر با استفاده از معیارهای مبتنی بر تئوری اطلاعات، میزان اهمیت هر ویژگی را محاسبه می‌کنند. این روش‌ها از سرعت بالایی، برخوردار و برای داده‌های با ابعاد بالا مناسب هستند. در روش پوششی، معیار انتخاب ویژگی، کارایی الگوریتم طبقه‌بندی است و بنابراین این روش از دقت بالایی در مقایسه با روش‌های فیلتر برخوردار است، اما به دلیل اجرای الگوریتم طبقه‌بند برای ارزیابی میزان شایستگی هر زیرمجموعه از ویژگی‌های انتخاب‌شده، سرعت کار روش پوششی کمتر از روش فیلتر است. به همین دلیل استفاده از روش‌های پوششی

چکیده: طبقه‌بندی چندبرچسبی یکی از مسائل مهم در یادگیری ماشین است که کارایی الگوریتم‌های این طبقه‌بندی با افزایش ابعاد مسأله به شدت کاهش می‌یابد. انتخاب ویژگی، یکی از راهکارهای اصلی برای کاهش ابعاد در مسائل چندبرچسبی است. انتخاب ویژگی چندبرچسبی یک راهکار NP Hard است و تا کنون تعدادی راهکار مبتنی بر هوش جمعی و الگوریتم‌های تکاملی برای آن ارائه شده است. افزایش ابعاد مسأله منجر به افزایش فضای جستجو و به تبع، کاهش کارایی و همچنین کاهش سرعت همگرایی این الگوریتم‌ها می‌شود. در این مقاله یک راهکار هوش جمعی ترکیبی مبتنی بر الگوریتم دودویی بهینه‌سازی ازدحام ذرات و استراتژی جستجوی محلی برای انتخاب ویژگی چندبرچسبی ارائه شده است. برای افزایش سرعت همگرایی، در استراتژی جستجوی محلی، ویژگی‌ها بر اساس میزان افزونه‌بودن و میزان ارتباط با خروجی مسأله به دو دسته تقسیم‌بندی می‌شوند. دسته اول را ویژگی‌هایی تشکیل می‌دهند که شباهت زیادی به کلاس مسأله و شباهت کمتری به سایر ویژگی‌ها دارند و دسته دوم هم ویژگی‌های افزونه و کمتر مرتبط است. بر این اساس، یک اپراتور محلی به الگوریتم بهینه‌سازی ازدحام ذرات اضافه شده که منجر به کاهش ویژگی‌های غیر مرتبط و افزونه هر جواب می‌شود. اعمال این اپراتور منجر به افزایش سرعت همگرایی الگوریتم پیشنهادی در مقایسه با سایر الگوریتم‌های ارائه‌شده در این زمینه می‌شود. عملکرد روش پیشنهادی با شناخته‌شده‌ترین روش‌های انتخاب ویژگی، بر روی مجموعه داده‌های مختلف مقایسه گردیده است. نتایج آزمایش‌ها نشان دادند که روش پیشنهادی از نظر دقت، دارای عملکردی مناسب است.

کلیدواژه: انتخاب ویژگی، طبقه‌بندی چندبرچسبی، استراتژی جستجوی محلی، هوش جمعی، بهینه‌سازی ازدحام ذرات.

## ۱- مقدمه

یادگیری بانظارت به استنتاج روابط بین نمونه‌های ورودی و برچسب‌های کلاس می‌پردازد. در الگوریتم‌های کلاسیک یادگیری ماشین، به هر نمونه یک برچسب تخصیص داده می‌شود، در حالی که در بسیاری از سناریوهای دنیای واقعی، یک نمونه ممکن است با چندین برچسب مرتبط باشد. برای مثال در طبقه‌بندی اخبار، یک قسمت از اخبار

این مقاله در تاریخ ۲۸ آبان ماه ۱۴۰۰ دریافت و در تاریخ ۱۶ خرداد ماه ۱۴۰۱ بازنگری شد.

آذر رفیعی، دانشکده مهندسی کامپیوتر، واحد سنج، دانشگاه آزاد اسلامی، سنج، ایران، (email: azar.rafi@gmail.com).

پرهام مرادی (نویسنده مسئول)، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنج، ایران، (email: p.moradi@uok.ac.ir).

عبدالباقی قادرزاده، دانشکده مهندسی کامپیوتر، واحد سنج، دانشگاه آزاد اسلامی، سنج، ایران، (email: b.ghaderzadeh@iausdj.ac.ir).

چندبرچسبی بیان می‌شود. مقایسه کارایی این روش با سایر روش‌های انتخاب ویژگی چندبرچسبی در بخش چهارم آمده و نهایتاً در بخش پنجم به جمع‌بندی کلی از روش ارائه‌شده پرداخته می‌شود.

## ۲- پیشینه تحقیق

هدف از انتخاب ویژگی، کاهش ابعاد فضای ویژگی و بهبود کارایی و عملکرد طبقه‌بندی از طریق از بین بردن ویژگی‌های نامربوط و تکراری است. به طور کلی، روش‌های انتخاب ویژگی چندبرچسبی به سه دسته فیلتر، پیچشی<sup>۱</sup> و ترکیبی تقسیم می‌شوند. روش‌های فیلتر مستقل از الگوریتم‌های یادگیری، ابعاد داده‌ها را نیز کاهش می‌دهند. این روش‌ها ویژگی‌های موجود را بر مبنای تعدادی معیار، رتبه‌بندی نموده و ویژگی‌هایی را که دارای رتبه کافی نیستند، حذف می‌کنند. تا کنون روش‌های فیلتر بسیاری مانند ReliefF و مشتقات آن [۳]، بهره اطلاعاتی<sup>۲</sup> [۴] و اطلاعات متقابل<sup>۳</sup> [۲۹] در زمینه انتخاب ویژگی مطرح شده است.

روش‌های انتخاب ویژگی چندبرچسبی از نظر مدل داده به دو دسته کلی تقسیم می‌شوند: تبدیل مسأله و انطباق الگوریتم. در الگوریتم‌های تبدیل مسأله، ابتدا داده‌های چندبرچسبی به یک یا چند داده تک‌برچسبی تبدیل می‌گردند و سپس انتخاب ویژگی روی داده تک‌برچسبی انجام می‌شود. Chen و همکارانش [۳۰]، یک روش تبدیل بر پایه آنتروپی پیشنهاد می‌دهند. همچنین Spolaor و همکارانش [۳۱]، LP و BR را برای تبدیل داده استفاده می‌کنند و برای انتخاب ویژگی، روش‌های ReliefF و بهره اطلاعاتی را به کار می‌گیرند و نهایتاً کارایی این روش‌ها را با یکدیگر مقایسه می‌کنند. قابل ذکر است که روش BR استفاده‌شده در این کار، قادر به تشخیص همبستگی بین برچسب‌ها نیست و LP نیز در فاز آموزش دارای پیچیدگی بالایی می‌باشد. Relief تعداد ویژگی‌های کمتری را انتخاب می‌کند و به این دلیل است که تعامل بین ویژگی‌ها را در نظر می‌گیرد. این روش، یک جستجوی چندمتغیره بوده و قادر به تمیزدادن دو نمونه از یک کلاس یا کلاس‌های متفاوت است. کاهش ابعاد بدون لطمه‌زدن به عملکرد طبقه‌بندی، وابستگی بین هر ویژگی و برچسب کلاس را مشخص می‌کند.

در [۳۲]، یک روش تبدیل مسأله هرس‌شده برای تبدیل داده چندبرچسبی به تک‌برچسبی به کار گرفته شده و سپس یک روش انتخاب ویژگی حریصانه بر پایه اطلاعات چندبعدی اجرا گردیده است. این روش در ابعاد بالا کارایی چندانی ندارد و در آن، LP قادر به تشخیص همبستگی بین برچسب‌ها است و وابستگی بین برچسب‌ها در نظر گرفته می‌شود. همچنین روشی در [۳۳]، نتایج اولیه [۳۲] را گسترش و روشی برای انتخاب خودکار پارامتر هرس پیشنهاد می‌دهد. روشی نیز در [۳] پیشنهاد گردیده که مسأله چندبرچسبی را با استفاده از تبدیل هرس‌شده (PPT) به مسأله تک‌برچسبی تبدیل می‌کند و سپس الگوریتم ReliefF برای اختصاص وزن به هر ویژگی به کار گرفته می‌شود. این روش نشان می‌دهد که روش‌های انطباقی بهتر از روش‌های تبدیل مسأله هستند. در [۳۴] یک روش انتخاب ویژگی سریع مبتنی بر رتبه‌دهی ویژگی ارائه شده که با حذف محاسبات غیر ضروری و شناسایی ترکیب برچسب‌های مهم

برای داده‌های با ابعاد بالا محدودیت ایجاد می‌کند. روش‌های ترکیبی با به کارگیری مزیت‌های هر ۲ روش، قادر به تولید جواب‌هایی بهتر هستند. در سال‌های اخیر، الگوریتم‌های هوش جمعی و تکاملی زیادی از جمله بهینه‌سازی کلونی مورچه (ACO) [۶]، [۷] و [۱۵] تا [۱۸]، الگوریتم ژنتیک (GA) [۱۱] و [۱۹] تا [۲۱]، تیرید شبیه‌سازی شده (SA) [۲۲] و [۲۳] و بهینه‌سازی ازدحام ذرات (PSO) [۱۰] و [۲۴] تا [۲۷] برای انتخاب ویژگی تک‌برچسبه ارائه شده‌اند.

به تازگی در [۱۰] یک راهکار پوششی انتخاب ویژگی چندبرچسبه بر اساس الگوریتم ازدحام ذرات آمده است. نتایج ارائه‌شده در این مقاله نشان از کارایی این الگوریتم دارند اما در بعضی شرایط، نتایج ارائه‌شده نشان‌دهنده دقت پایین و گیرافتادن در بهینه محلی هستند. برای رفع این مشکل در این مقاله یک راهکار جستجوی محلی ارائه گردیده که نه تنها منجر به بهبود کارایی و دقت الگوریتم می‌شود، بلکه سرعت همگرایی را افزایش می‌دهد. بر اساس این استراتژی، ابتدا میزان شباهت هر ویژگی با سایر ویژگی‌ها بر اساس تئوری اطلاعات متقابل محاسبه می‌شود. ویژگی‌های افزونه به طور معمول، شباهت بیشتری با سایر ویژگی‌ها دارند. همچنین میزان ارتباط هر ویژگی با برچسب‌های مسأله محاسبه می‌شود و نهایتاً بر اساس این دو پارامتر، معیاری ارائه می‌گردد که میزان مرتبط‌بودن و همچنین افزونه‌بودن یک ویژگی را نشان می‌دهد. ویژگی‌ها را بر اساس این معیار، مرتب و به ۲ دسته تقسیم می‌کنیم. ویژگی‌هایی که مستقل و مرتبط هستند، در دسته اول قرار می‌گیرند و ویژگی‌هایی که دارای ارتباط کمتر و افزونگی بیشتر هستند، در دسته دوم قرار دارند. سپس یک مرحله جهش به الگوریتم بهینه‌سازی ازدحام ذرات اضافه شده و بر اساس این مرحله برای هر جواب در جمعیت، ویژگی‌های مرتبط را از دسته اول به جواب اضافه کرده و ویژگی‌هایی را که در دسته دوم قرار دارند، از جواب حذف می‌کنیم. این استراتژی منجر به افزایش سرعت همگرایی و خروج از بهینه محلی خواهد شد.

به عبارت دیگر در این روش از همبستگی بین ویژگی‌ها و برچسب‌های مسأله برای بهتر شدن فرایند جستجو استفاده می‌شود. به علاوه، این روش با استفاده از تخمین اندازه زیرمجموعه، سعی در کاهش چشم‌گیر تعداد ویژگی‌های زیرمجموعه نهایی دارد. اهداف و نوآوری‌های اصلی پژوهش در زیر به صورت خلاصه ارائه شده است:

- از یک استراتژی جستجوی محلی که با الگوریتم بهینه‌سازی ذرات ادغام شده است، به منظور کاهش ابعاد زیرمجموعه ویژگی و انتخاب زیرمجموعه مطلوب استفاده می‌گردد.
- هدف استراتژی جستجوی محلی، هدایت فرایند جستجوی الگوریتم بهینه‌سازی ازدحام ذرات به منظور انتخاب ویژگی‌های متمایز با در نظر گرفتن اطلاعات همبستگی آنها است.
- استفاده از اطلاعات همبستگی برای هدایت فرایند جستجو در الگوریتم ازدحام ذرات به گونه‌ای است که ویژگی‌های غیر هم‌بسته (غیر مشابه) با احتمال بیشتری از ویژگی‌های هم‌بسته (مشابه) انتخاب می‌شوند.
- انتخاب تعداد کم ویژگی‌های برجسته با استفاده از یک طرح تعیین اندازه زیرمجموعه خاص انجام می‌شود.
- این الگوریتم در مقایسه با الگوریتم [۱۰] کارایی بهتر و سرعت همگرایی بالاتری دارد.

مقاله حاضر در ۵ بخش نگارش گردیده است که در بخش دوم کارهای انجام‌شده در زمینه انتخاب ویژگی داده‌های چندبرچسبی مرور خواهند گردید. در بخش سوم، روش پیشنهادی برای انتخاب ویژگی داده‌های

1. Filter
2. Wrapper
3. Gain Information
4. Mutual Information

جدول ۱: خلاصه‌ای از کارهای انجام‌شده در زمینه انتخاب ویژگی چندبرچسبی روش‌ها.

روش‌ها	انتخاب ویژگی	تبدیل داده	روش جستجو	الگوریتم طبقه‌بندی	مجموعه داده
[۲۹] ELA	فیلتر	Entropy	بهره اطلاعاتی - OCFS	SVM	متن
[۳۶] MGFS	فیلتر	-	PageRank	MLKNN	دامنه‌های مختلف
[۱۸] MLACO	فیلتر	-	الگوریتم کلونی مورچه	MLKNN	دامنه‌های مختلف
RF-BR, RF-LP, [۳۰] IG-BR, IG-LP	فیلتر	LP, BR	اطلاعات متقابل - ReliefF	BRKNN	دامنه‌های مختلف
[۳۲] MI	فیلتر	PPT	اطلاعات متقابل	MLKNN	بیولوژی، متن، تصویر
[۸] Memetic	پیچشی	-	الگوریتم ژنتیک	MLNB	دامنه‌های مختلف
[۹] MLFS	پیچشی	-	اطلاعات متقابل - الگوریتم ژنتیک	MLKNN	متن، تصویر
[۱۳] MEFS	ترکیبی	LP	Max Avg, LP-Chi	LEAD, SVM, MLNB, MLKNN	موزیک
[۳۷] Sun	فیلتر	-	اطلاعات متقابل	MLKNN	دامنه‌های مختلف
[۱۱] MLNB	پیچشی	-	الگوریتم ژنتیک	MLNB	بیولوژی، تصویر
[۲۹] MI	پیچشی	-	اطلاعات متقابل	MLNB	بیولوژی، متن، تصویر
[۳] ReliefF	فیلتر	PPT	ReliefF	MLKNN, BRKNN	دامنه‌های مختلف
[۱۲] Khan	پیچشی	-	الگوریتم ژنتیک	SVM, DT	موزیک
[۱۰] MPFS	پیچشی	-	الگوریتم ازدحام ذرات	MLKNN	دامنه‌های مختلف
[۳۵] MMFS	پیچشی	-	NSGA-II	MLKNN	موزیک، بیولوژی، متن
[۳۶] MLInfoGain	فیلتر	-	بهره اطلاعاتی	Various Classifiers	دامنه‌های مختلف
[۲] MFNMI	فیلتر	-	اطلاعات متقابل	MLKNN	دامنه‌های مختلف
[۵] MDMR	فیلتر	-	حداقل افزونگی - حداکثر ارتباط	MLKNN, LIFT	تصویر
[۱۴] MLMLFS	ترکیبی	-	$l_2, p$ -norm	MLKNN	دامنه‌های مختلف
[۳۱] Doquire	فیلتر	PPT, LP	اطلاعات متقابل	MLKNN, SVM	دامنه‌های مختلف
[۴] IGMF	فیلتر	-	بهره اطلاعاتی	SVM, MLKNN	بیولوژی، تصویر، موزیک

می‌شود [۸]. همچنین در [۳۵]، یک الگوریتم انتخاب ویژگی چندبرچسبی بر پایه اطلاعات متقابل و الگوریتم ژنتیک آمده است. در مرحله اول، اطلاعات متقابل برای تکمیل ویژگی‌های انتخابی به صورت محلی به کار گرفته شده و سپس الگوریتم ژنتیک طبق نتایج این مرحله، زیرمجموعه ویژگی‌های بهینه سراسری را انتخاب می‌کند.

یک سری از کارها نیز در زمینه انتخاب ویژگی چندبرچسبی انجام شده که از مفهوم بهینه‌سازی چندهدفه استفاده می‌کنند. همه این روش‌ها جزء روش‌های پیچشی هستند که به طور معمول بر اساس متاهیوریستیک‌ها می‌باشند. برای مثال، ژانگ و همکارانش در [۱۰] از الگوریتم ازدحام ذرات و [۱۲] از الگوریتم ژنتیک استفاده می‌کنند. در همه این کارها، دو یا سه معیار ارزیابی به طور هم‌زمان بهینه می‌شوند. مثلاً در [۱۲] از ۳ معیار ارزیابی دقت، افت همینگ<sup>۲</sup> و micro-Average و در [۱۰] از افت همینگ و دقت استفاده می‌شود. در پژوهشی دیگر و در ابتدای کار، فضای راه حل به یک گراف وزن‌دار کامل، تبدیل و سپس با استفاده از روش PageRank، اهمیت هر ویژگی محاسبه می‌شود [۳۶]. در [۱۸]، یک روش فیلتر بر پایه الگوریتم کلونی مورچه پیشنهاد گردیده که شباهت کسینوسی نرمال بین ویژگی‌ها و برچسب‌ها در نظر گرفته می‌شود.

جدول ۱ کارهای انجام‌شده را در زمینه انتخاب ویژگی چندبرچسبی نشان می‌دهد. ستون گروه انتخاب ویژگی مختص به نوع دسته‌بندی انتخاب ویژگی (فیلتر، پیچشی و ترکیبی) است. قسمت روش تبدیل داده نشان می‌دهد که آیا در این روش تکنیک تبدیل استفاده شده یا خیر؟ و در صورت استفاده، روش تبدیل ذکر گردیده است. در ستون روش جستجو، تکنیک انتخاب ویژگی مورد استفاده در هر روش آمده است. ستون

جهت انتخاب ویژگی چندبرچسبی سریع، سرعت فرایند جستجو را افزایش می‌دهد. Lin و همکارانش در پژوهش دیگری، یک رویکرد انتخاب ویژگی چندبرچسبی که ویژگی‌های مطلوب را بر پایه اطلاعات متقابل همسایگی انتخاب می‌کند، پیشنهاد داده‌اند. در ابتدا همه نمونه‌ها تحت برچسب‌های مختلف با استفاده از حاشیه نمونه‌ها دانه‌بندی می‌گردند و ۳ اطلاعات متقابل همسایگی مختلف برای یادگیر چندبرچسبی تعریف می‌شود. سپس آنها یک تابع چندهدفه بهینه را برای اندازه‌گیری کیفیت ویژگی‌های منتخب معرفی کرده‌اند [۲].

در روش‌های ترکیبی، جستجو برای یافتن زیرمجموعه مناسبی از ویژگی‌ها توسط یک الگوریتم یادگیری انجام می‌شود. You و همکارانش در پژوهشی، یک روش انتخاب ویژگی ترکیبی را ارائه می‌دهند که یک معیار ارزیابی به نام ریسک پیش‌بینی برای ارزیابی ویژگی‌ها اتخاذ و از استراتژی جستجوی عقب‌گرد برای یافتن زیرمجموعه ویژگی مناسب استفاده می‌کند [۱۳]. همچنین در [۱۴]، یک روش انتخاب ویژگی ترکیبی برای طبقه‌بندی چندبرچسبی با برچسب‌های مفقود پیشنهاد شده است. برچسب‌های گم‌شده با استفاده از رگرسیون خطی پوشش داده شده‌اند و ویژگی‌های منتخب توسط  $p$ -norm و مؤثر  $l_2$  ( $0 < p \leq 1$ ) مشخص می‌شوند. در این کار پیچیدگی محاسباتی طبقه‌بند بالاست و استقلال بین برچسب‌ها در نظر گرفته نشده است.

روش‌های پیچشی از یک الگوریتم یادگیری از پیش تعریف شده استفاده می‌کنند و این روش، ویژگی‌هایی را انتخاب می‌کند که کارایی الگوریتم یادگیری را افزایش دهند. در [۱۱]، استخراج ویژگی بر پایه PCA است که با یک روش انتخاب ویژگی پیچشی پیاده‌سازی شده با الگوریتم ژنتیک ترکیب گردیده است. در پژوهشی دیگر، یک روش انتخاب ویژگی چندبرچسبی مبتنی بر روش متامیک برای تعریف مجدد زیرمجموعه ویژگی‌های یافته‌شده توسط جستجوی ژنتیک استفاده

1. Accuracy  
2. Hamming Loss

بر این ضعف‌ها، روش انتخاب ویژگی جدیدی مبتنی بر PSO پیشنهاد شده که در آن، زیرمجموعه نهایی ویژگی‌ها از طریق چند مرحله انتخاب می‌گردد. ابتدا اندازه مجموعه ویژگی به صورت خودکار تخمین زده شده و سپس تمامی ویژگی‌ها به گروه‌های مشابه و غیر مشابه با استفاده از اطلاعات همبستگی آنها تقسیم می‌شوند. در اینجا روش PSO باینری با یک استراتژی جستجوی محلی خاص که اطلاعات محلی ویژگی‌ها را درون فرایند جستجو در نظر می‌گیرد، ترکیب شده و سپس تعدادی از ذرات از پیش تعیین شده تولید می‌گردند. این ذرات با توجه به بهترین موقعیت محلی خود و بهترین سراسری جمعیت به سمت موقعیت‌های جدید حرکت می‌کنند. هر ذره، درون موقعیت محلی خود را با توجه به اطلاعات همبستگی ویژگی‌ها جستجو می‌کند. در این مرحله، تناسب هر ذره محاسبه شده و در آخر، بهترین ذره‌های سراسری و محلی جایگزین مقادیر پیشین می‌شوند. هنگامی که به معیار توقف رسیدند، مجموعه ویژگی نهایی اعلام می‌شود. شبه‌کد این روش در شکل ۱ آمده است.

### ۳-۱ تعیین اندازه زیرمجموعه و گروه‌بندی ویژگی‌ها

برای تعیین اندازه زیرمجموعه از یک تابع تصادفی احتمال استفاده می‌شود که تلاش می‌کند تعداد کمتری از ویژگی‌ها را برای تعیین اندازه زیرمجموعه  $s$  فراهم کند [۴۰]. ابتدا از (۱) برای تعیین اندازه زیرمجموعه ویژگی‌ها استفاده می‌شود [۴۱]

$$l_s = \frac{f - s}{\sum_{i=1}^f (f - i)} \quad (1)$$

که  $f$  تعداد ویژگی‌های اصلی،  $s$  تعداد ویژگی‌های انتخاب شده،  $l$  تفاوت بین  $s$  و  $f$  و  $l_s$  مقدار احتمال تخمین  $s$  به عنوان یک تعداد اولیه از ویژگی‌ها می‌باشد. واضح است در صورتی که  $l_s$  بیشینه باشد،  $s$  کمینه می‌شود. تعداد اولیه ویژگی‌ها ( $s$ ) با استفاده از چرخه رولت<sup>۲</sup> و بر پایه احتمال  $l_s$  به دست می‌آید.  $s$  به صورت تصادفی در محدوده  $[N1, N2]$  انتخاب می‌شود، در حالی که شروع  $s$  برابر  $N1$  و  $N2 = e.f$  است و  $e$  یک پارامتر تعریف شده توسط کاربر است که برای کنترل  $N2$  استفاده می‌شود. اگر  $e$  به ۱ نزدیکتر باشد،  $N2$  به  $f$  نزدیکتر می‌شود و بنابراین فضای جستجو برای یافتن ویژگی‌های برجسته بزرگتر می‌شود و بالطبع، هزینه محاسباتی بالایی خواهد داشت و همچنین ممکن است زیرمجموعه‌های ویژگی بی‌اثری تولید شود.

لازم به ذکر است که تلاش طرح بر این است که اندازه زیرمجموعه کوچک‌تری را ارائه دهد و اندازه زیر مجموعه به مقدار  $e$  که توسط کاربر تعیین می‌شود، بستگی دارد. سپس ویژگی‌ها به دو گروه مشابه و غیر مشابه تقسیم می‌شوند. هدف از این کار، یافتن روابط بین ویژگی‌ها است به طوری که این الگوریتم می‌تواند ویژگی‌های متمایزی را برای مدل‌های یادگیری قوی انتخاب کند [۴۲]. همبستگی از رایج‌ترین و مفیدترین معیارهایی است که میزان روابط بین دو متغیر را توصیف می‌کند. در ابتدا برای اندازه‌گیری همبستگی بین ویژگی‌های مختلف یک مجموعه، ضریب همبستگی پیرسون از (۲) به دست می‌آید [۴۳]

$$CC_{ij} = \frac{\sum_{k=1}^m (x_i(k) - \bar{x}_i)(x_j(k) - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_i(k) - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_j(k) - \bar{x}_j)^2}} \quad (2)$$

### Algorithm 1: MPSOLS<sup>1</sup>-Multi label Feature Selection based on Particle Swarm Optimization and Local search strategy

**Input:** Feature =  $\{f_1, f_2, \dots, f_D\}$ ,  $NC_{max}$ : maximum cycle number, NP: number of particles.

**Output:** Final set of features =  $\{f_1, f_2, \dots, f_k\}$ .

**Begin** algorithm

**Begin** initialization

Determining number of feature (s) to k using Eq (1)

**for** i = 1 to NP **do**

    Create random particle  $X_{ij}$ .

    Create random velocity  $V_{ij}$ .

**end for**

**Compute**  $CC_{ij}$  and  $Cor_i$  for all features using Eqs (2), (3)

$Cor_i \geq Cor_{mid}$  Similar feature

$Cor_i < Cor_{mid}$  Dissimilar feature

**end** initialization

**for** i = 1 to  $NC_{max}$  **do**

**Compute** updating the particle velocity using Eq (4)

**Compute** updating the position of particle using Eq (6)

$X_s$  = Similar features.

$X_d$  = Dissimilar features.

**Remove** all feature in  $X_s$  that is 0 in particle x.

**Remove** all feature in  $X_d$  that is 0 in particle x.

**Calculate** the value of  $n_s$  and  $n_d$

**Perform** "particle movement" on each position of particle and replace it.

$f_i = fitness(x'_i)$

**if**  $f_i > fitness(x_i^{best})$  **then**

$x_i^{best} = x'_i$

**if**  $f_i > fitness(x_g^{best})$  **then**

$x_g^{best} = x'_i$

**end for**

**Return**  $x_g^{best}$

**end** algorithm

شکل ۱: شبه‌کد روش پیشنهادی MPSOLS.

الگوریتم طبقه‌بندی، استراتژی طبقه‌بندی و نهایتاً ستون مجموعه‌های داده، لیستی از مجموعه‌های داده استفاده شده در هر روش را نشان می‌دهد.

### ۳- روش پیشنهادی

الگوریتم PSO یک روش بهینه‌سازی مبتنی بر قواعد احتمال است که در سال ۱۹۹۵ برای اولین بار با الهام از رفتار پرندگان در هنگام جستجوی غذا توسط کندی و آبرهات ابداع شد [۳۸]. در این الگوریتم، ابتدا یک مجموعه جواب اولیه تولید می‌شود و سپس برای یافتن جواب بهینه در فضای پاسخ‌های ممکن، یا به‌هنگام کردن نسل‌ها، جستجوی پاسخ انجام می‌گیرد. هر ذره به صورت چندبعدی با دو مقدار موقعیت و سرعت تعریف می‌گردد و در هر مرحله از حرکت ذره، با دو شاخص سرعت و موقعیت، بهترین پاسخ‌ها از لحاظ شایستگی برای تمام ذرات تعیین می‌شوند [۳۹].

یافتن بهترین راه حل در زمینه انتخاب ویژگی با توجه به استراتژی جستجوی سراسری، از مهم‌ترین مزایای الگوریتم ازدحام ذرات (PSO) است. همچنین از معایب این الگوریتم می‌توان به همگرایی زودرس و ضعف در تنظیم دقیق نزدیک نقاط بهینه محلی اشاره کرد. برای غلبه

اگر  $V_i(t+1)$  بزرگ‌تر از مقدار تصادفی بود، مقدار موقعیت با ۱ نمایش داده می‌شود و به این معنی است که ویژگی متناظر در به‌روزرسانی بعدی نیز انتخاب می‌شود. اما اگر  $V_i(t+1)$  کوچک‌تر از مقدار تصادفی بود، مقدار موقعیت با صفر نمایش داده می‌شود و به این معنی است که ویژگی متناظر در به‌روزرسانی بعدی نیز انتخاب نشده است.

### ۳-۳ استراتژی جستجوی محلی

در این گام برای یک ذره، عملگر "ADD" و "Delete" به منظور بهبود جستجوی محلی ذره به کار گرفته شده است. یک ذره برای انتخاب تعدادی از ویژگی‌های مطلوب، عملگر "ADD" را به کار می‌برد و عملگر "Delete" جهت حذف یک تعداد از ویژگی‌های موجود استفاده می‌شود. به عبارتی اپراتور "ADD"، ویژگی‌های غیر مشابه را به ذره اضافه می‌کند و ویژگی‌های مشابه با اپراتور "Delete" از ذره حذف می‌شوند [۴۰]. در عملگر جستجوی محلی، ابتدا همه ویژگی‌های انتخاب شده توسط ذره استخراج می‌گردند. سپس الگوریتم در هر ذره، تعدادی از بیت‌های یک را در ذره تولیدشده جدید مشخص می‌کند (مثال ۱۰۱۱۰۰۰۱۱۱) و آنها را درون زیرگروه قرار می‌دهد (مثال:  $X = \{f_1, f_7, f_7, f_8, f_8, f_8, f_8\}$ ).

هر مؤلفه  $X$  با  $D$  و  $S$  مقایسه و زیرگروه  $X$  به  $X_s$  و  $X_d$  تقسیم می‌شود.  $X_d$  شامل ویژگی‌هایی است که در  $D$  وجود دارند و ویژگی‌های باقیمانده  $X$  در  $X_s$  قرار می‌گیرند [۴۴]. سپس در گام بعدی، همه ویژگی‌های  $X_s$  و  $X_d$  به ترتیب صعودی بودن و بر اساس مقدار همبستگی آنها مرتب می‌شوند که اولین ویژگی و آخرین ویژگی از  $X_d$  و  $X_s$  دارای بیشترین تمایز و بیشترین شباهت هستند. در حرکت ذره، نیازمند کنترل بادقت تعداد بیت‌های ۱ در ذره تولیدشده جدید هستیم و با توجه به به‌روزرسانی سرعت ذرات، اگر سرعت از مقدار rand مشخص شده بیشتر باشد، موقعیت ذره مورد نظر ۱ و اگر کمتر باشد، صفر خواهد شد. در پایان، تعدادی از ویژگی‌های مشابه و غیر مشابه با محاسبه مقدار  $n_s$  و  $n_d$  معین می‌شوند. در اینجا  $n_s = \mu \cdot s$  و  $n_d = (1 - \mu) \cdot s$  است، در حالی که  $\mu$  یک پارامتر خاص کاربر و  $s$  اندازه زیرمجموعه اولیه ویژگی‌هاست که در مرحله اول تخمین زده شد. سپس طول  $X_s$  و  $X_d$  را با یکدیگر مقایسه می‌کنیم، به طوری که اگر تعداد ویژگی‌های غیر مشابه در ذره کوچک‌تر از  $n_d$  باشد ( $|X_d| < n_d$ )، آن گاه  $n_d - |X_d|$  ویژگی درون  $D - X_d$  به ذره اضافه می‌شود و در غیر این صورت،  $|X_d| - n_d$  ویژگی درون  $X_d$  باید از ذره حذف گردد. این در حالی است که اگر تعداد ویژگی‌های مشابه بزرگ‌تر از  $n_s$  باشد ( $|X_s| > n_s$ )، باید  $|X_s| - n_s$  ویژگی مشابه درون  $X_s$  از ذره حذف شوند. به عبارت دیگر، زمانی که تعداد ویژگی‌های مشابه در ذره تولیدشده کوچک‌تر از  $n_s$  است،  $|X_s| - n_s$  ویژگی درون  $S - X_s$  به ذره اضافه می‌شود.

### ۳-۴ محاسبه تناسب

روش پیشنهادی از ML-kNN برای ارزیابی یک راه حل زیرمجموعه ویژگی کاندیدا استفاده می‌کند. قبل از فرایند ارزیابی، ابتدا هر ویژگی با مقیاس‌هایی بین -۱ و ۱ نرمال می‌شود. فرایند نرمال‌سازی، ویژگی‌های غالب با مقادیر عددی بیشتر را به ویژگی‌هایی با محدوده عددی محدود، تغییر می‌دهد. نتایج آزمایش‌ها نشان داد که مقیاس‌بندی مقادیر ویژگی‌ها می‌تواند به بهبود دقت طبقه‌بندی کمک کند. بنابراین در این مقاله، یک روش نرمال‌سازی خطی برای مقیاس‌بندی مجموعه‌های داده به شرح زیر استفاده می‌شود

که  $CC_{ij}$  ضریب همبستگی بین دو ویژگی  $i$  و  $j$ ،  $m$  تعداد نمونه‌ها و  $x_i(k)$  و  $x_j(k)$  نشان‌دهنده مقادیر بردارهای ویژگی  $i$  و  $j$  برای نمونه  $k$ ام است. اگر مقدار ضریب همبستگی بالاتر باشد، به این معنی است که دو ویژگی شباهت بالایی به یکدیگر دارند. از سوی دیگر، مقادیر کمتر نشان‌دهنده تشابه پایین هستند. پس از محاسبه ضریب همبستگی برای همه ترکیب‌های ممکن از ویژگی‌ها، مقدار همبستگی برای ویژگی  $i$  به صورت زیر به دست می‌آید

$$Cor_i = \frac{\sum_{j=1}^f |CC_{ij}|}{f-1} \quad \text{if } i \neq j \quad (3)$$

که  $f$  تعداد کل ویژگی‌ها و  $CC_{ij}$  ضریب همبستگی بین دو ویژگی  $i$  و  $j$  است. نهایتاً دو گروه به اندازه‌های  $f/2$  ایجاد می‌شوند که گروه اول شامل ویژگی‌هایی با بالاترین میزان همبستگی است که آن را گروه مشابه ( $S$ ) می‌نامیم و گروه دوم، باقیمانده ویژگی‌ها با شباهت کمتر است که آن را گروه غیر مشابه ( $D$ ) می‌نامیم. بدین ترتیب ویژگی اول در گروه  $S$  و آخرین ویژگی در گروه  $D$  به ترتیب بیشترین و کمترین ویژگی‌های همبسته در مجموعه داده هستند.

### ۳-۲ مقداردهی و به‌روزرسانی ذرات

در این مرحله، هر ذره با یک بردار باینری تعریف می‌شود که طول آن برابر تعداد ویژگی‌های اصلی است. اگر مقدار یک سلول از بردار صفر باشد، به این معنی است که ویژگی متناظر انتخاب نشده و اگر یک باشد، به این معنی است که ویژگی متناظر انتخاب شده است. سپس برای هر ذره، یک بردار سرعت به طولی برابر طول بردارهای ذره با استفاده از یک مولد float تصادفی تولید می‌شود. سرعت هر ذره  $i$  مطابق با (۴) تغییر خواهد کرد

$$V_i(t+1) = V_i(t) + C_1 \cdot rand_{,1} \cdot (pbest_i(t) - X_i(t)) + C_2 \cdot rand_{,2} \cdot (gbest_i(t) - X_i(t)) \quad (4)$$

که در آن  $t$  نمایانگر تعداد تکرار،  $pbest_i$  بهترین مقدار به دست آمده توسط ذره  $i$ ام از آغاز اجرای الگوریتم تا تکرار  $t$ ام،  $gbest_i$  بهترین مقدار به دست آمده توسط جمعیت از آغاز اجرای الگوریتم تا تکرار  $t$ ام و  $C_1$  و  $C_2$  دو عدد حقیقی ثابت هستند. همچنین  $r_1$  و  $r_2$  بردارهایی تصادفی هستند که مقادیر آنها در بازه صفر و یک قرار دارند. لازم به ذکر است که اگر سرعت ذرات بالا برود و از  $V_{max}$  تجاوز کند، باید سرعت این بُعد را به  $V_{max}$  مطابق معادله زیر محدود کرد

$$\text{if } V_i(t+1) \notin (V_{min}, V_{max}) \text{ then } V_i(t+1) = \max(\min(V_{max}, V_i(t+1)), V_{min}) \quad (5)$$

در حالی که  $V_{min}$  و  $V_{max}$  پارامترهای خاص کاربر هستند (در اینجا  $V_{min} = -4$  و  $V_{max} = 4$  در نظر گرفته شده است). باید توجه داشت که موقعیت ذره در هر تکرار و با تغییر سرعت به‌روزرسانی می‌گردد که در اینجا موقعیت هر ذره با تابع  $(V_i(t+1))$  بر اساس فرمول زیر به‌روزرسانی می‌شود

$$S(V_i(t+1)) = \frac{1}{1 + e^{-V_i}} \quad \text{if } rand < S(V_i(t+1)) \text{ then } X_i(t+1) = 1 \quad \text{else } X_i(t+1) = 0 \quad (6)$$

یک خطا! این معیار تعداد دفعاتی را می‌شمرد که برچسب با بیشترین درجه تعلق به یک نمونه، متعلق به آن نمونه نباشد

$$One - error = \frac{1}{p} \sum_{i=1}^p [\arg \max_{y \in \gamma} f(x_i, y)] \notin Y_i \quad (9)$$

پوشش! این معیار تعداد گام‌هایی را می‌شمرد که باید لیست برچسب‌های درجه‌بندی شده را طی نمود تا تمام برچسب‌های هر نمونه را پوشش داد

$$Coverage = \frac{1}{p} \sum_{i=1}^p \max_{y \in \gamma} rank_f(x_i, y) - 1 \quad (10)$$

افت رتبه! میانگین تعداد دفعاتی را محاسبه می‌کند که یک برچسب مرتبط، رتبه بالاتری از یک برچسب نامربوط گرفته است

$$Ranking - loss = \frac{1}{p} \sum_{i=1}^p \frac{1}{|Y_i|} \times \left\{ \frac{1}{|Y_i|} \sum_{y \in Y_i} \left[ f(x_i, y') \leq f(x_i, y) \right] \right\} \quad (11)$$

میانگین صحت! میانگین درصد برچسب‌های مرتبط را از یک نمونه که رتبه آنها از رتبه یک برچسب مرتبط خاص متعلق به آن نمونه بیشتر باشد، می‌شمرد

$$Average\ precision = \frac{1}{p} \times \sum_{i=1}^p \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{\left[ f(x_i, y') \leq f(x_i, y) \right]}{rank_f(x_i, y)} \quad (12)$$

### ۴-۳ نتایج و بحث

برای ارزیابی کارایی روش پیشنهادی و مقایسه آن با روش‌های MPSOFS [۱۰]، MLNB [۱۱]، ParetoFS [۴۷]، PMU [۴۸]، ELA-CHI [۳۰]، PPT-CHI [۳۵]، PPT-MI [۳۲] و MLACO [۱۸] آزمایش‌هایی انجام شده و این در حالی است که در تمامی روش‌ها از طبقه‌بند ML-KNN [۴۶] استفاده گردیده است. همچنین به جز روش MPSOFS و MLNB که پیچشی هستند، سایر روش‌ها از نوع فیلتر می‌باشند. در آزمایش‌ها از مجموعه‌های داده‌ای مانند Yeast، Scene و Birds استفاده شده است. جداول ۳ تا ۵، مقایسه عملکرد الگوریتم‌های مورد آزمایش را از منظر معیارهای افت همینگ، افت رتبه، یک-خطا، پوشش و میانگین صحت بر روی مجموعه‌های مختلف داده نشان می‌دهد. برای هر معیار ارزیابی، علامت "↑" به صورت "مقدار بیشتر، بهتر است" و علامت "↓" به صورت "مقدار کمتر، بهتر است" تعریف می‌شود و همچنین نتیجه بهتر به صورت پررنگ‌تر نشان داده شده است. در این جدول، آزمون ویلکاکسون برای مقایسه کارایی روش‌های انتخاب ویژگی مورد استفاده قرار می‌گیرد. این روش، یک آزمون آماری استنباطی است که برای ارزیابی همانندی دو نمونه وابسته با مقیاس رتبه‌ای به کار می‌رود. این روش تفاوت بین هر جفت داده با ارزش  $p$  را محاسبه کرده

2. One Error
3. Coverage
4. Ranking Loss
5. Average Precision
6. Wilcoxon

جدول ۲: مشخصات ۶ مجموعه داده با کاربردهای مختلف.

اسم	تعداد برچسب	تعداد ویژگی	تعداد نمونه	دامنه
Yeast	۱۴	۱۰۳	۲۴۱۷	زیست‌شناسی
Birds	۱۹	۲۶۰	۶۴۵	صوت
Scene	۶	۲۹۴	۲۴۰۷	تصویر

$$X^{new} = l + [(u - l) \times \frac{x - x_{min}}{x_{max} - x_{min}}] \quad (7)$$

که در آن  $l$  و  $u$  به ترتیب کران پایین و بالای فرایند نرمال‌سازی و  $x_{min}$  و  $x_{max}$  حداکثر و حداقل مقادیر ویژگی  $x$  هستند. پس از فرایند نرمال‌سازی، مجموعه داده جدید از مجموعه داده اصلی با ویژگی‌هایی که در راه حل ذره وجود داشتند، استخراج گردیده و سپس مجموعه داده‌ها به بخش آموزش و تست تقسیم می‌شوند (۶۰٪ برای مجموعه آموزشی و ۴۰٪ برای مجموعه تست). در این مرحله، روش اعتبارسنجی متقابل (۱۰-fold) [۴۵] برای ارزیابی هر ذره با استفاده از طبقه‌بندی کننده ML-kNN به کار گرفته می‌شود. پس برای هر مجموعه داده، مجموعه آموزشی به ۱۰ قسمت تقسیم می‌گردد که هر بخش همان نسبت را برای هر کلاس از داده‌ها به اشتراک می‌گذارد. بنابراین برای هر مجموعه داده، ۹ قسمت اول داده‌ها در فرایند آموزش برای ساخت مدل یادگیری استفاده می‌شود، در حالی که آخرین بخش در فرایند اعتبارسنجی برای ارزیابی مقدار تناسب ذره مورد استفاده قرار می‌گیرد. اگر دقت دو راه حل یکسان باشد، راه حل با استفاده از تعداد کوچک‌تر ویژگی‌ها انتخاب می‌شود.

### ۴- نتایج آزمایش‌ها

#### ۴-۱ مجموعه‌های داده

در این مقاله، ۳ مجموعه داده از کاربردهای مختلف از پایگاه داده مولان تهیه شده که جدول ۲ مشخصه این مجموعه‌های داده شامل اسم، تعداد برچسب، تعداد ویژگی، تعداد نمونه‌ها و دامنه آنها را نشان می‌دهد.

#### ۴-۲ معیارهای ارزیابی

برای ارزیابی کارایی طبقه‌بندی یک طبقه‌بند چندبرچسبی، معیارهای مختلفی طراحی شده است (مانند افت همینگ، یک-خطا، پوشش، افت رتبه، میانگین صحت، دقت و ... [۴۶]). در این مقاله نیز از تعدادی از این معیارها برای ارزیابی این روش و مقایسه آن با روش‌های دیگر انتخاب ویژگی چندبرچسبی استفاده شده است. فرضاً  $T = \{(t_i, Y_i), i = 1, \dots, p\}$  یک مجموعه آزمون باشد که  $Y_i \subseteq L$ ، زیرمجموعه واقعی و  $Z_i \subseteq L$ ، زیرمجموعه پیش‌بینی شده متناظر با نمونه  $t_i$  باشد. همچنین فرض کنید که  $f(x, y)$  امتیازی است که به برچسب  $y$  برای نمونه  $x$  داده می‌شود. این معیارها در (۸) تا (۱۲) معرفی شده‌اند.

افت همینگ: معیار افت همینگ، درصد برچسب‌هایی را که به درستی طبقه‌بندی نشده‌اند، محاسبه می‌کند. کمترین مقدار، بهترین کارایی را دارد

$$Hamming - loss = \frac{1}{P} \sum_{i=1}^p \frac{|Y_i \Delta Z_i|}{|L|} \quad (8)$$

$\Delta$  تفاوت متقارن بین دو مجموعه است.

جدول ۳: نتایج مقایسه روش پیشنهادی با سایر روش‌ها با ۱۰۰ تکرار مستقل روی مجموعه داده BIRDS.

ویژگی‌کامون	MPSOLS	ParetoFS	PMU	ELA-CHI	PPT-MI	PPT-CHI	MLNB	MLACO	MPSOFS	الگوریتم
+	۰٫۲۸۹	۰٫۱۲۴۷۸۵	۰٫۱۰۱۴۵۲	۰٫۱۴۵۹۶۸	۰٫۱۹۰۳۰۲	۰٫۱۰۹۰۳۸	۰٫۰۹۳۰۴۲	۰٫۰۹۸۴۵۵	۰٫۰۴۳۶۷	افت همینگ ↓
+	۰٫۲۰۵۴۶	۰٫۲۸۷۹۳۳	۰٫۲۴۸۹۷	۰٫۲۶۳۶۱	۰٫۲۶۶۹۸	۰٫۲۱۸۵۷	۰٫۲۲۲۵۲۲	۰٫۲۲۷۲۳۴	۰٫۲۱۱۶۲۸	افت رتبه ↓
+	۰٫۵۵۱۲۵	۰٫۵۷۵۴۸	۰٫۶۰۲۸۴۵	۰٫۶۱۳۴۲	۰٫۵۸۹۰۰۵	۰٫۶۰۱۴	۰٫۵۹۳۰۲۳	۰٫۵۱۰۱۵۳	۰٫۵۱۷۴۴۲	یک-خطا ↓
+	۲٫۵۰۹۸۳	۲٫۵۶۲۴۵۶	۲٫۵۲۵۷	۲٫۶۶۵۷۸	۲٫۹۴۷۵۶	۲٫۷۵۱۴۲	۲٫۶۶۲۵۳۹	۳٫۵۱۵۶۰۱	۲٫۵۵۷۲۷۶	پوشش ↓
+	۰٫۵۵۷۹۶	۰٫۵۴۷۰۲۹	۰٫۵۳۵۲	۰٫۵۱۰۲۲	۰٫۵۱۴۲۶	۰٫۵۲۴۶۵	۰٫۴۸۷۶۸۶	۰٫۵۰۷۶۲۳	۰٫۵۳۴۸۱۳	میانگین صحت ↑

جدول ۴: نتایج مقایسه روش پیشنهادی با سایر روش‌ها با ۱۰۰ تکرار مستقل روی مجموعه داده YEAST.

ویژگی‌کامون	MPSOLS	ParetoFS	PMU	ELA-CHI	PPT-MI	PPT-CHI	MLNB	MLACO	MPSOFS	الگوریتم
+	۰٫۱۷۰۰۱	۰٫۵۰۴۵۴۱	۰٫۳۴۴۷۸۵	۰٫۳۷۸۹۵۴	۰٫۴۱۵۴۷۸	۰٫۳۵۵۷۰۰	۰٫۲۱۴۰۵۲	۰٫۲۱۴۷۷۴	۰٫۱۹۳۷۲۲	افت همینگ ↓
+	۰٫۱۰۵۵۴۶	۰٫۷۶۰۴۵	۰٫۷۰۲۴۵	۰٫۶۷۳۲۱۴	۰٫۷۳۱۲	۰٫۷۰۶۸۴	۰٫۱۸۱۴۶۷	۰٫۱۹۲۸۵۹	۰٫۱۶۹۲۲۵	افت رتبه ↓
+	۰٫۲۴۲۳۳	۰٫۳۵۵۳	۰٫۹۰۳۷	۰٫۸۱۲۰۳۴	۰٫۶۸۲۷۱	۰٫۸۳۷۱۰	۰٫۲۴۵۳۶۵	۰٫۲۷۴۵۶۱	۰٫۲۴۸۶۳۷	یک-خطا ↓
+	۶٫۴۴۸۲۶	۷٫۱۹۶۲۰۵	۷٫۲۳۳۴۷	۷٫۲۲۷۳۹۸	۷٫۲۱۹۳۸۷	۷٫۲۳۰۹۴۵	۶٫۵۸۹۶۶۷	۶٫۷۵۴۸۱۲	۶٫۳۴۸۹۶۴	پوشش ↓
+	۰٫۷۷۰۰۲	۰٫۲۳۶۴۷۸	۰٫۲۵۵۸۹۷	۰٫۲۵۳۱۴۵	۰٫۲۴۲۲۴۴	۰٫۲۵۵۵۹۷	۰٫۲۴۴۱۸۶	۰٫۲۷۲۷۳۵	۰٫۲۵۴۸۱۳	میانگین صحت ↑

جدول ۵: نتایج مقایسه روش پیشنهادی با سایر روش‌ها با ۱۰۰ تکرار مستقل روی مجموعه داده SCENE.

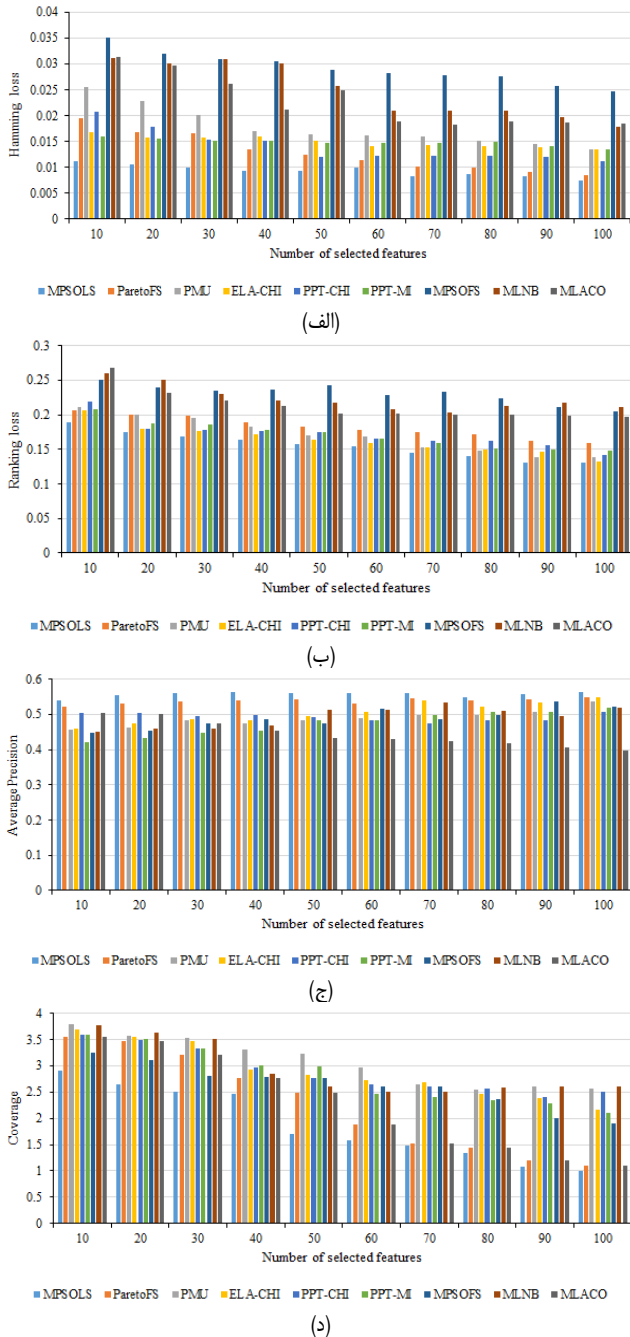
ویژگی‌کامون	MPSOLS	ParetoFS	PMU	ELA-CHI	PPT-MI	PPT-CHI	MLNB	MLACO	MPSOFS	الگوریتم
+	۰٫۰۷۹۴۲۱	۰٫۳۵۸۱۱۰	۰٫۳۵۰۵۶۴	۰٫۲۵۴۴۸۷	۰٫۳۳۶۴۲۵	۰٫۲۶۱۱۵۸	۰٫۱۳۳۶۴	۰٫۱۳۵۰۰۱	۰٫۰۸۷۳۷۵	افت همینگ ↓
+	۰٫۱۰۱۸۱	۰٫۷۴۱۲۵	۰٫۵۳۴۲۳	۰٫۵۶۰۲۴	۰٫۶۱۲۴۵	۰٫۵۶۰۲	۰٫۱۲۳۲۴۴	۰٫۲۴۰۳۶۹	۰٫۰۹۴۲۷۳	افت رتبه ↓
+	۰٫۲۴۴۶۱	۰٫۸۶۷۸۵	۱٫۹۲۹۰۶	۱٫۸۲۴۶۱	۱٫۳۹۴۴۹	۱٫۷۴۲۵۴	۰٫۳۱۶۸۹	۰٫۴۵۷۴۸	۰٫۲۳۱۶۰۵	یک-خطا ↓
+	۰٫۶۰۰۱۳	۱٫۱۴۳۵۰	۱٫۱۷۸۴۷۷	۱٫۱۷۷۴۳۱	۱٫۱۶۹۵۰۲	۱٫۱۷۶۲۴	۰٫۷۱۴۸۸۳	۰٫۸۷۸۶۲۵	۰٫۵۸۱۹۴	پوشش ↓
+	۰٫۸۳۹۸۸	۰٫۴۲۹۹۸	۰٫۶۷۸۹۵	۰٫۶۶۴۵۸	۰٫۵۶۸۵۴	۰٫۶۵۸۲۴۱	۰٫۸۰۳۴۹۱	۰٫۷۶۵۴۸	۰٫۸۲۳۴۷	میانگین صحت ↑

میانگین صحت مشاهده می‌کنیم که اختلاف روش پیشنهادی در مجموعه داده Yeast نسبت به روش ParetoFS، ۰٫۵۳ می‌باشد و این در حالی است که با توجه به این معیار، در سایر مجموعه‌های داده نیز روش پیشنهادی نتایج بهتری را کسب کرده است. روش پیشنهادی از نظر معیار افت همینگ در برخی مجموعه‌های داده با اختلاف بالایی نسبت به سایر روش‌ها جایگاه اول را به دست آورده است. برای نمونه، اختلاف این معیار در مجموعه داده Birds با توجه به روش پیشنهادی نسبت به روش PPT-MI، ۰٫۱۵ یا در مجموعه داده Scene نسبت به روش ParetoFS، ۰٫۲۷ می‌باشد. روش پیشنهادی در تمامی معیارها در مقایسه با روش MLACO نیز نتایج قابل توجهی را به دست آورده است. جدول ۳ نشان می‌دهد که روش پیشنهادی بهترین عملکرد را در بین تمام روش‌های انتخاب ویژگی دارد. همان‌طور که مشاهده می‌شود، در جدول ۳ نتایج آماری به دست آمده کارایی بهتر روش پیشنهادی نسبت به دیگر روش‌ها را بیان می‌کنند.

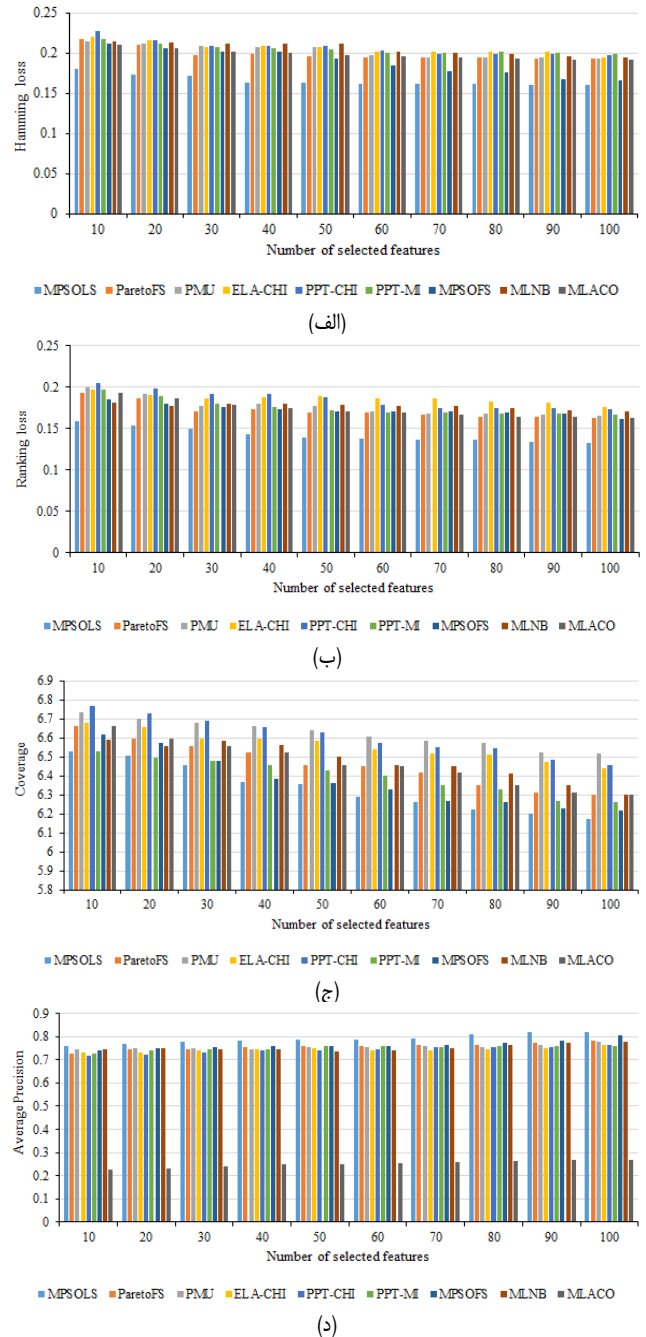
در شکل‌های ۲ تا ۴ به ازای هر یک از مجموعه‌های داده، تمامی معیارها بررسی شده است. محور افقی و عمودی در این شکل‌ها به ترتیب نشان‌دهنده تعداد ویژگی‌های انتخاب شده و معیار ارزیابی هستند. در هر کدام از شکل‌ها، هشت نمودار به رنگ‌های مختلف آمده که هر کدام نشان‌دهنده یک روش انتخاب ویژگی چندبرچسبی است. این روش‌ها به

و تفاوت‌ها را تجزیه و تحلیل می‌کند. در مورد مقایسه روش‌های انتخاب ویژگی، فرضیه صفر نشان می‌دهد که هیچ تفاوتی در مورد عملکرد دو روش انتخاب ویژگی وجود ندارد. اگر مقدار  $p$  کمتر یا برابر با یک سطح معنی‌دار مشخص باشد ( $\alpha = 0.05$ )، فرضیه صفر رد می‌شود و می‌توان استنباط کرد که تفاوت قابل توجهی بین دو روش وجود دارد [۴۹]. آخرین ستون از هر جدول، نتیجه مقایسه آماری روش پیشنهادی با سایر روش‌ها است. علامت مثبت نشان‌دهنده برتری روش پیشنهادی در مقابل دیگر روش‌های انتخاب ویژگی است و علامت منفی نشان می‌دهد که روش پیشنهادی برتر نیست، در حالی که علامت = بیان می‌کند که هیچ تفاوت قابل توجهی بین عملکرد دو روش انتخاب ویژگی وجود ندارد.

همان‌طور که مشاهده می‌کنید در تمام معیارها به جز پوشش، روش پیشنهادی بهترین عملکرد را داشته و در معیار پوشش نیز روش پیشنهادی پس از روش MPSOFS در جایگاه دوم قرار دارد. همچنین در بعضی از مجموعه‌های داده، اختلاف روش پیشنهادی با سایر روش‌ها بسیار قابل توجه است. به عنوان مثال، پوشش در مجموعه داده‌ای Scene برای روش پیشنهادی برابر ۰٫۶۰۰۱۳ و اختلاف پوشش در این روش با روش PPT-CHI تقریباً برابر ۰٫۵۷ است. همچنین معیار افت رتبه در مجموعه داده Birds برای روش پیشنهادی برابر ۰٫۲۰۵۴ است که اختلاف آن با روش ParetoFS، ۰٫۱۰۸ و با روش ELA-CHI، ۰٫۱۰۵ می‌باشد. در معیار



شکل ۳: نتایج آزمایش‌ها بر روی مجموعه داده Birds، (الف) افت همینگ، (ب) افت رتبه، (ج) پوشش و (د) میانگین صحت.



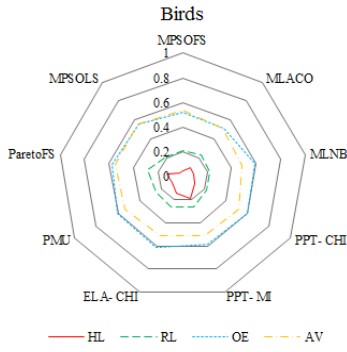
شکل ۴: نتایج آزمایش‌ها بر روی مجموعه داده Yeast، (الف) افت همینگ، (ب) افت رتبه، (ج) پوشش و (د) میانگین صحت.

طی می‌کند و این بدان معنی است که روش پیشنهادی ما تأثیر بسزایی بر انتخاب ویژگی‌های مطلوب دارد. همچنین در این شکل‌ها مشاهده می‌کنیم که این روش، بیشترین میانگین صحت را در تمام مجموعه‌های داده به دست آورده است. این معیار در مجموعه داده Yeast بیشترین شیب صعودی را با توجه به افزایش تعداد ویژگی‌های انتخاب شده طی کرده است. شکل‌های ۲ تا ۴ به طور واضح نشان می‌دهند که روش پیشنهادی، زیرمجموعه‌های ویژگی بهتری را نسبت به روش‌های دیگر برای تمام معیارهای ارزیابی تولید می‌کند.

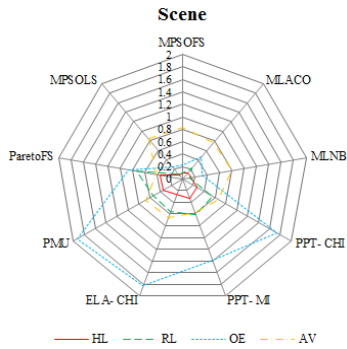
برای هر یک از مجموعه‌های داده با توجه به معیارهای ارزیابی افت همینگ، افت رتبه، یک-خطا و میانگین صحت، نمودارهای راداری رسم شده است. در شکل‌های ۵ تا ۷ هرچه فاصله خطوط مشخص شده برای به معیار افت همینگ، افت رتبه و یک-خطا از مرکز کمتر باشد، آن روش کارآمدتر است و برای میانگین صحت، این فاصله هرچه از مرکز بیشتر

ترتیب، مقایسه عملکرد الگوریتم‌های مورد آزمایش را از نظر معیارهای افت همینگ، افت رتبه، یک-خطا، پوشش و میانگین صحت نشان می‌دهند. در همه نمودارهای موجود در آزمایش‌ها، محور افقی نشان‌دهنده تعداد ویژگی‌های انتخاب شده و محور عمودی بیانگر کارایی طبقه‌بندی است. تعداد ویژگی انتخاب شده توسط کاربر تعریف می‌شود و هر روش دارای ۲۰۰ تکرار مستقل می‌باشد. در شکل‌های ۲ تا ۴ واضح است که روش پیشنهادی، کمترین افت همینگ را در مجموعه‌های داده Yeast و Scene به دست آورده است. در این مجموعه‌های داده، روش پیشنهادی با افزایش تعداد ویژگی‌های انتخاب شده به نتایج بهتری دست یافته و می‌توان نتیجه گرفت که قدرت ویژگی‌های پیش‌بینی‌کننده به طور قابل توجهی افزایش یافته است. معیار پوشش در مجموعه‌های داده Birds و Yeast کمترین مقدار را نسبت به سایر روش‌ها دارد و این معیار در این مجموعه‌های داده با بیشتر شدن تعداد ویژگی‌ها، شیب نزولی بیشتری را

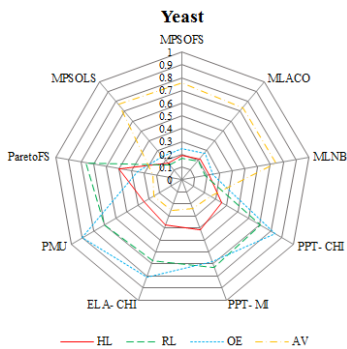




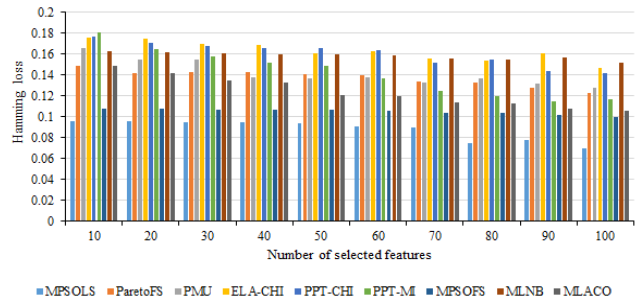
شکل ۵: نمودار راداری بر روی مجموعه داده Birds.



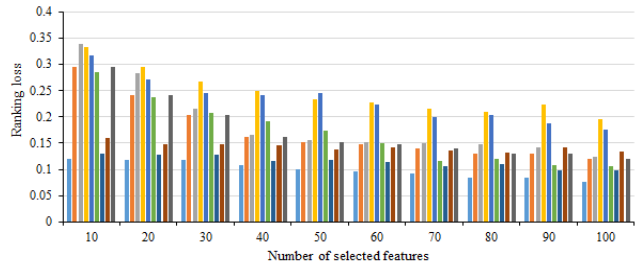
شکل ۶: نمودار راداری بر روی مجموعه داده Scene.



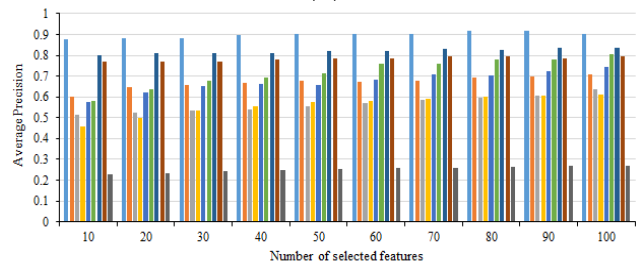
شکل ۷: نمودار راداری بر روی مجموعه داده Yeast.



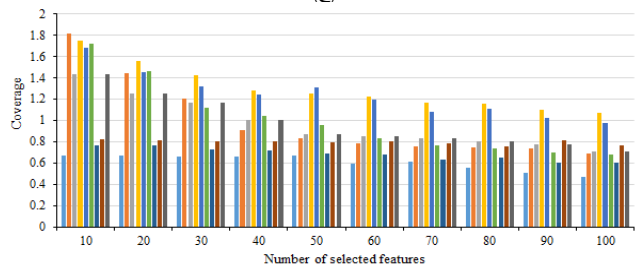
(الف)



(ب)



(ج)



(د)

شکل ۴: نتایج آزمایش‌ها بر روی مجموعه داده Scene، (الف) افت همینگ، (ب) افت رتبه، (ج) پوشش و (د) میانگین صحت.

باشد، نتایج بهتری را خواهیم داشت. در هر یک از این نمودارها می‌توان برتری روش پیشنهادی را به وضوح مشاهده نمود.

### ۵- نتیجه گیری

در این مقاله یک روش انتخاب ویژگی چندبرچسبی ترکیبی مبتنی بر هوش جمعی و به کارگیری یک استراتژی جدید جستجوی محلی پیشنهاد شده است. این روش اندازه ویژگی‌های مطلوب را با تخمین اندازه زیرمجموعه ویژگی‌ها کاهش می‌دهد. هدف از ارائه این الگوریتم، بهبود کارایی الگوریتم ازدحام ذرات برای مسائل مقیاس بزرگ در مسأله انتخاب ویژگی است. برای این منظور، یک اپراتور جدید به یکی از مراحل الگوریتم بهینه‌سازی ازدحام ذرات استفاده شد که کار این اپراتور، جستجوی محلی در فضای جواب جهت افزایش سرعت همگرایی و جلوگیری از گیرافتادن در بهینه محلی است. برای این کار ابتدا ویژگی‌ها

به دو دسته مرتبط و غیر مرتبط تقسیم‌بندی شدند. در هر مرحله، اپراتور جستجوی محلی سعی در اضافه‌نمودن ویژگی‌های مرتبط به هر عضو جمعیت (یک جواب مسأله) و حذف ویژگی‌های غیر مرتبط از آن دارد. برای تشخیص مرتبط و نامرتب بودن ویژگی‌ها از تئوری مبتنی بر اطلاعات متقابل و برای ارزیابی راهکار پیشنهادی از معیارهای ارزیابی مبتنی بر طبقه‌بندی چندبرچسبی استفاده شده است. با توجه به این که الگوریتم جستجو ویژگی تصادفی دارد، برای اثبات معتبر بودن نتایج به دست آمده، از آزمون‌های آماری همچون تست تی و تست ویلکاکسون استفاده شده است. در این کار ما از معیارهای ارزیابی مختلف و شش مجموعه داده چندبرچسبی برای مقایسه روش پیشنهادی با روش‌های دیگر استفاده کرده‌ایم. نتایج آزمایش‌ها نشان داد که در بیشتر نمونه‌ها، روش پیشنهادی کارایی بهتری نسبت به روش‌های دیگر انتخاب ویژگی چندبرچسبی دارد. ایده‌های مختلفی برای پیشبرد راهکار ارائه شده این مقاله در آینده وجود دارد. یکی از این ایده‌ها می‌تواند استفاده از الگوریتم‌های خوشه‌بندی برای تقسیم‌بندی ویژگی‌ها به دسته‌های مختلف به جای استفاده از دو دسته باشد. این کار می‌تواند منجر به افزایش بیشتر سرعت همگرایی و دقت بیشتر الگوریتم شود. زیرا از هر خوشه سعی می‌شود یک ویژگی مرتبط انتخاب شود و از انتخاب ویژگی‌های افزونه

- [20] M. Rostami and P. Moradi, "A clustering based genetic algorithm for feature selection," in *Proc. 6th Conf. on Information and Knowledge Technology, IKT'14*, pp. 112-116, Shahrood, Iran, 27-29 May. 2014.
- [21] T. M. Hamdani, J. M. Won, A. M. Alimi, and F. Karray, "Hierarchical genetic algorithm with new evaluation function and bi-coded representation for the selection of features considering their confidence rate," *Applied Soft Computing*, vol. 11, no. 2, pp. 2501-2509, Mar. 2011.
- [22] S. W. Lin, Z. J. Lee, S. C. Chen, and T. Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Applied Soft Computing*, vol. 8, no. 4, pp. 1505-1512, Sep. 2008.
- [23] S. W. Lin, T. Y. Tseng, S. Y. Chou, and S. C. Chen, "A simulated-annealing-based approach for simultaneous parameter optimization and feature selection of back-propagation networks," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1491-1499, Feb. 2008.
- [24] L. Y. Chuang, C. H. Yang, and J. C. Li, "Chaotic maps based on binary particle swarm optimization for feature selection," *Applied Soft Computing*, vol. 11, no. 1, pp. 239-248, Jan. 2011.
- [25] Y. Liu, et al., "An improved particle swarm optimization for feature selection," *J. of Bionic Engineering*, vol. 8, no. 2, pp. 191-200, Jun. 2011.
- [26] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: novel initialisation and updating mechanisms," *Applied Soft Computing*, vol. 18, pp. 261-276, May 2014.
- [27] H. M. Abdelsalam and A. M. Mohamed, "Optimal sequencing of design projects' activities using discrete particle swarm optimisation," *International J. of Bio-Inspired Computation*, vol. 4, no. 2, pp. 100-110, 2012.
- [28] K. Demir, B. H. Nguyen, B. Xue, and M. Zhang, "Particle swarm optimisation for sparsity-based feature selection in multi-label classification," in *Proc. of the Genetic and Evolutionary Computation Conf. Companion*, pp. 232-235, Boston, MA, USA, 9-13 Jul. 2022.
- [29] J. Lee and D. W. Kim, "Mutual information-based multi-label feature selection using interaction information," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2013-2025, Mar. 2015.
- [30] W. Chen, J. Yan, B. Zhang, Z. Chen, and Q. Yang, "Document transformation for multi-label feature selection in text categorization," in *Proc of 7th IEEE Int. Conf. on Data Mining, ICDM'07*, vol. ???, pp. 451-456, Omaha, NE, USA, 28-31 Oct. 2007.
- [31] N. Spolaor, E. A. Cherman, M. C. Monard, and H. D. Lee, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electronic Notes in Theoretical Computer Science*, vol. 292, pp. 135-151, Mar. 2013.
- [32] G. Doquire and M. Verleysen, "Feature selection for multi-label classification problems," in *Proc of Int. Work-Confer. on Artificial Neural Networks, IWANN'11*, pp. 9-16, Torremolinos-Málaga, Spain, 8-10 Jun. 2011.
- [33] G. Doquire and M. Verleysen, "Mutual information-based feature selection for multilabel classification," *Neurocomputing*, vol. 122, pp. 148-155, Dec. 2013.
- [34] J. Lee and D. W. Kim, "Fast multi-label feature selection based on information-theoretic feature ranking," *Pattern Recognition*, vol. 48, no. 9, pp. 2761-2771, Sept. 2015.
- [35] J. Read, B. Pfahringer, and G. Holmes, "Multi-label classification using ensembles of pruned sets," in *Proc of 8th IEEE Int. Conf. on Data Mining*, pp. 995-1000, Pisa, Italy, 15-19 Dec. 2008.
- [36] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MGFS: a multi-label graph-based feature selection algorithm via PageRank centrality," *Expert Systems with Applications*, vol. 142, Article ID: 113024, Mar. 2020.
- [37] Z. Sun, et al., "Mutual information based multi-label feature selection via constrained convex optimization," *Neurocomputing*, vol. 329, pp. 447-456, Feb. 2019.
- [38] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of Int. Conf. on Neural Networks, ICNN'95*, vol. 4, pp. 1942-1948, Perth, Australia, 27 Nov.-1 Dec. 1995.
- [39] ح. افراخته و پ. مرادی، "روشی جدید بمنظور خوشه‌بندی داده‌های سرعت باد در نیروگاه‌های بادی با استفاده از الگوریتم‌های PSO و FCM"، *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، ب- مهندسی کامپیوتر، سال ۸، شماره ۳، صص. ۲۱۴-۲۱۰، پاییز ۱۳۸۹.
- [40] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, no. 17, pp. 2914-2928, Oct. 2011.
- [41] D. P. Muni, N. R. Pal, and J. Das, *Genetic Programming for Simultaneous Feature Selection and Classifier Design*, 2006.
- اجتناب می‌گردد. ایده دیگر، گسترش الگوریتم برای مسائل انتخاب ویژگی در جریان است. در این مسائل ویژگی‌ها ثابت نیستند و در هر مرحله ممکن است ویژگی‌های جدیدی ارائه شود. الگوریتم هوش جمعی و استراتژی جستجوی محلی این مقاله می‌تواند برای این نوع مسائل تطبیق داده شود.
- ## مرجع
- [۱] م. رحمانی‌نیا و پ. مرادی، "یک الگوریتم انتخاب ویژگی برخط در جریان داده‌ها با استفاده از اطلاعات متقابل چندمتغیره"، *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، ب- مهندسی کامپیوتر، سال ۱۸، شماره ۴، صص. ۳۳۶-۳۲۷، زمستان ۱۳۹۹.
- [2] Y. Lin, Q. Hu, J. Liu, J. Chen, and J. Duan, "Multi-label feature selection based on neighborhood mutual information," *Applied Soft Computing*, vol. 38, pp. 244-256, Jan. 2016.
- [3] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, pp. 168-182, Aug. 2015.
- [4] L. Li, et al., "Multi-label feature selection via information gain," in *Proc. Int. Conf. on Advanced Data Mining and Applications, ADMA'14*, pp. 345-355, Guilin, China, 19-21 Dec. 2014.
- [5] Y. Lin, Q. Hu, J. Liu, and J. Duan, "Multi-label feature selection based on max-dependency and min-redundancy," *Neurocomputing*, vol. 168, pp. 92-103, Nov. 2015.
- [6] S. Tabakhi and P. Moradi, "Relevance-redundancy feature selection based on ant colony optimization," *Pattern Recognition*, vol. 48, no. 9, pp. 2798-2811, Sept. 2015.
- [7] P. Moradi and M. Rostami, "Integration of graph clustering with ant colony optimization for feature selection," *Knowledge-Based Systems*, vol. 84, pp. 144-161, Aug. 2015.
- [8] J. Lee and D. W. Kim, "Memetic feature selection algorithm for multi-label classification," *Information Sciences*, vol. 293, pp. 80-96, Feb. 2015.
- [9] Y. Yu and Y. Wang, "Feature selection for multi-label learning using mutual information and GA," in *Proc. 9th Int. Conf. on Rough Sets and Knowledge Technology, RSKT'14*, pp. 454-463, Shanghai, China, 24-26 Oct. 2014.
- [10] Y. Zhang, D. W. Gong, X. Y. Sun, and Y. N. Guo, "A PSO-based multi-objective multi-label feature selection method in classification," *Scientific Reports*, vol. 7, Article ID: 376, Mar. 2017.
- [11] M. L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive bayes classification," *Information Sciences*, vol. 179, no. 19, pp. 3218-3229, Sept. 2009.
- [12] M. A. Khan, A. Ekbal, E. L. Mencía, and J. Fürnkranz, "Multi-objective optimisation-based feature selection for multi-label classification," in *Proc. Int. Conf. on Applications of Natural Language to Information Systems, NLDB'17*, pp. 38-41, Liege, Belgium, 21-23 Jun. 2017.
- [13] M. You, J. Liu, G. Z. Li, and Y. Chen, "Embedded feature selection for multi-label classification of music emotions," *International J. of Computational Intelligence Systems*, vol. 5, no. 4, pp. 668-678, Aug. 2012.
- [14] P. Zhu, Q. Xu, Q. Hu, C. Zhang, and H. Zhao, "Multi-label feature selection with missing labels," *Pattern Recognition*, vol. 74, pp. 488-502, Feb. 2018.
- [15] S. Tabakhi, A. Najafi, R. Ranjbar, and P. Moradi, "Gene selection for microarray data classification using a novel ant colony optimization," *Neurocomputing*, vol. 168, pp. 1024-1036, Nov. 2015.
- [16] R. K. Sivagaminathan and S. Ramakrishnan, "A hybrid approach for feature subset selection using neural networks and ant colony optimization," *Expert Systems with Applications*, vol. 33, no. 1, pp. 49-60, Jul. 2007.
- [17] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Text feature selection using ant colony optimization," *Expert Systems with Applications*, vol. 36, no. 3, pt. 2, pp. 6843-6853, Apr. 2009.
- [18] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MLACO: a multi-label feature selection algorithm based on ant colony optimization," *Knowledge-Based Systems*, vol. 192, Article ID: 105285, Mar. 2020.
- [19] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems*, vol. 13, no. 2, pp. 117-136, Mar. 1998.

**پرهام مرادی** دوره کارشناسی را در رشته مهندسی کامپیوتر در سال ۱۳۸۱ در دانشگاه صنعتی امیرکبیر به پایان رسانید. ایشان همچنین دوره کارشناسی ارشد و دکتری را در رشته علوم کامپیوتر دانشگاه صنعتی امیرکبیر به ترتیب در سال‌های ۱۳۸۴ و ۱۳۸۹ به پایان رسانید. ایشان بخشی از تحقیقات دوره دکتری را به عنوان فرصت مطالعاتی را در دانشگاه پلی تکنیک فدرال لوزان سوئیس (EPFL) در سال ۱۳۸۸ به انجام رسانید. نامبرده در سال ۱۳۸۹ به عنوان هیأت علمی در گروه مهندسی کامپیوتر دانشگاه کردستان مشغول فعالیت گردید. در ادامه در سال ۱۳۹۵ به مرتبه دانشیاری ارتقاء یافت. حاصل تحقیقات ایشان تاکنون انتشار بیش از ۹۰ مقاله در مجلات و کنفرانس‌های معتبر بوده است. ایشان همچنین بر اساس گزارش منتشره دانشگاه استنفورد، به عنوان دانشمند دو درصد برتر در سال ۲۰۱۹ شناخته شدند. تحقیقات ایشان شامل یادگیری ماشین، یادگیری تقویتی، تحلیل شبکه‌های اجتماعی، علوم داده و سیستم‌های توصیه‌گر است. ایشان در دانشگاه کردستان مسئولیت‌هایی همچون مدیر گروه، مدیر آموزش‌های آزاد و مجازی و ریاست دانشکده مهندسی را بر عهده داشته است.

**عبدالباقی قادرزاده** در سال ۱۳۸۳ مدرک کارشناسی خود را در رشته علوم کامپیوتر از دانشگاه تبریز و در سال ۱۳۸۶ مدرک کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات، طراحی و تولید نرم افزار از دانشگاه علم و صنعت ایران دریافت کرد. در سال ۱۳۸۷ به عنوان هیأت علمی ماده ۱ در دانشگاه علوم پزشکی کردستان مشغول به کار گردید. از سال ۱۳۹۰ به عنوان هیأت علمی در دانشگاه آزاد اسلامی واحد سنجش شروع به فعالیت کرد و از سال ۱۳۹۱ دوره دکتری مهندسی کامپیوتر سیستم‌های نرم‌افزاری را در دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران آغاز و در ۱۳۹۶ فارغ‌التحصیل گردید. زمینه‌های تحقیقاتی ایشان علوم داده، سیستم‌های توزیع‌شده و یادگیری برخط در این شاخه‌ها است.

- [42] M. M. Kabir, M. M. Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, pp. 3273-3283, Oct. 2010.
- [43] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proc. of the ACM Conf. on Computer Supported Cooperative Work, CSCW'94*, pp. 175-186, Chapel Hill, NC, USA, 22-26 Oct. 1994.
- [44] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. of the 18th Int. Conf. on Neural Information Processing Systems, NIPS'05*, pp. 507-514, Vancouver, Canada, 5-8 Dec. 2005.
- [45] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. of the Royal Statistical Society: Series B (Methodological)*, vol. 36, pp. 111-133, 1974.
- [46] M. L. Zhang and Z. H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038-2048, Jul. 2007.
- [47] S. Kashef and H. Nezamabadi-pour, "A label-specific multi-label feature selection algorithm based on the Pareto dominance concept," *Pattern Recognition*, vol. 88, pp. 654-667, 2019.
- [48] J. Lee and D. W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 349-357, Feb. 2013.
- [49] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed., Chapman & Hall, 2011.

**آذر رفیعی** کارشناسی و کارشناسی ارشد و دکتری مهندسی کامپیوتر را به ترتیب در سال‌های ۱۳۹۱ و ۱۳۹۴ و ۱۴۰۱ دریافت کرد. وی از سال ۱۳۹۴ در دانشگاه مشغول تدریس می‌باشد. علایق تحقیقاتی وی شامل داده‌کاوی، یادگیری ماشین و یادگیری عمیق می‌باشد.