

دانشگاه یزد

دانشکده مهندسی برق و کامپیوتر

پایان نامه

برای دریافت درجه کارشناسی ارشد

مهندسی کامپیوترگرایش نرم افزار

عنوان

## شرح گذاری خودکار مجموعه تصاویر در مقیاس بزرگ مبتنی بر انتخاب نمونه های اولیه

استاد راهنما

دکتر محمدعلی زارع چاهوکی

استاد مشاور

دکتر مهدی رضائیان

پژوهش و نگارش

حمید کارگر شورکی

اسفند ۱۳۹۴

دانشگاه اسلامی  
ایران

تعدیم به:

اسطوره‌های زندگیم،

پناه حشمتکنیم،

و امید بودنیم.

سپاس و ستایش خدای را که آثار قدرت او بر چهره روز روشن، تماباں است و انوار حکمت او در دل شب تار، در فدان. آفریدگاری که خویشتن را به ما شناساند و درهای علم را برمأکشود و عمری و فرصتی عطا فرمود تا مابداں، بنده ضعیف خویش را در طریق علم و معرفت بیازماید. سپاس، آن دو فرشته ای که از خواسته هاشان گذشتند، سختی ها را به جان خریدند و خود را سپر بلای مشکلات و ناملایمات کردند تا من به جایگاهی که اکنون در آن ایستاده ام برسم.

بر خود لازم می دانم از استاد باحالات و شایسته، جناب آقای دکتر محمد علی زارع چاهوکی که دکمال سع صدر، با حسن خلق و فروتنی، از یچ کلی در این عرصه بر من درین تتمودند و زحمت راهنمایی این پایان نامه را بر عهده گرفتند مشکر و قدردانی نمایم. بی شک بدون مساعدت ایشان، این پایان نامه به نتیجه مطلوب نمی رسید. برای ایشان آرزوی سلامتی، موفقیت و سر بلندی را دارم. از استاد گر اتفاقر، جناب آقای دکتر محمدی رضائیان نیز که زحمت مشاوره این پایان نامه را مستقبل شدند کمال مشکر و قدردانی را دارم.

همچنین از استاد گر اتفاقر، جناب آقای دکتر محمد تقی صادقی و جناب آقای دکتر علی محمد لطیف، که زحمت داوری این پایان نامه را مستقبل شدند، مشکر و قدردانی می کنم.

## چکیده

شرح‌گذاری خودکار تصاویر به معنی بکارگیری مدل‌های یادگیری برای توصیف محتوای دیداری تصاویر با استفاده از توصیفگرهای متنی است. مدل‌های شرح‌گذاری با تعداد زیاد تصاویر و کلمه‌های شرح‌گذاری در مجموعه‌ها، با چالش‌های جدیدی مواجه می‌شوند. از مهمترین این چالش‌ها مقیاس‌پذیری سیستم در رابطه با زمان شرح‌گذاری تصاویر و دقت شرح‌گذاری می‌باشد. در این پایان‌نامه به منظور حل چالش‌های مطرح شده در سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ از ایده انتخاب نمونه‌های اولیه بهره برده شده است.

در بکارگیری روش‌های انتخاب نمونه‌های اولیه در کاربرد شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ، دو رویکرد ارائه شده است. در رویکرد اول، ایده حذف نمونه‌های پرت طبقه‌ای در روش تنکسازی نزدیکترین همسایه (CNN) بکار برده شده و نسخه توسعه یافته‌ای از این روش ارائه شده است. شرح‌گذاری، مسئله طبقه‌بندی چندبرچسبی است. در رویکرد دوم ارائه شده، به منظور بهبود عملکرد و تطابق روش‌های انتخاب نمونه‌های اولیه در کاربردهای چندبرچسبی، روشی جدید از روش پایه ENN ارائه شده است. در روش MLENN-GKELM ارائه شده، تصمیم‌گیری در مورد حذف یک نمونه در دو مرحله صورت می‌گیرد. در ابتدا بردار برچسب نمونه کاندیدای حذف از روی نزدیکترین همسایه‌ها و با استفاده از دو الگوریتم پیش‌بینی حریصانه و KELM بدست می‌آید. سپس با محاسبه عدم شباهت دودویی دو بردار برچسب پیش‌بینی شده و واقعی، تصمیم‌گیری در مورد حذف نمونه صورت می‌گیرد. در این قسمت اثربخشی بکارگیری روابط محاسبه عدم شباهت دودویی متعدد مورد بررسی قرار گرفته است. این دو مرحله عمدۀ سهم رویکرد دوم ارائه شده هستند. با بکارگیری این دو عملکرد انتخاب نمونه‌های اولیه در ENN پایه بهبود می‌یابد. در شرح‌گذاری تصاویر بدون برچسب نیز از دو روش KNN وزن‌دار و GKELM استفاده شده و دقت شرح‌گذاری بهبود یافته است.

عملکرد مناسب رویکردهای ارائه شده در انتخاب نمونه‌های موثر در فرآیند شرح‌گذاری و بهبود دقت شرح‌گذاری با انجام آزمایش‌های متعدد بر روی دادگان مقیاس بالا نشان داده شده است. برای این منظور از مجموعه تصاویر مقیاس بزرگ NUS-WIDE-، NUS-WIDE-Object-، NUS-WIDE- است.

NUS-WIDE و NUS-WIDE-Lite Scene استفاده شده است. نتایج در سه بخش نرخ کاهش نمونه‌ها، دقت شرح‌گذاری و تعادل مابین دقت شرح‌گذاری و نرخ کاهش ارائه شده است. نرخ کاهش بدست آمده در رویکردهای پیشنهادی در این چهار مجموعه به ترتیب برابر  $74/15\%$ ،  $60/68\%$ ،  $79/09\%$  و  $43/90\%$  بدست آمده است. همچنین دقت شرح‌گذاری در معیار F1 ماکرو نیز به ترتیب برابر  $36/31\%$ ،  $33/24\%$ ،  $25/63\%$  و  $22/72\%$  می‌باشد. در نهایت تعادل مابین دقت شرح‌گذاری و نرخ کاهش در این چهار مجموعه به ترتیب برابر  $92/62\%$ ،  $27/20\%$ ،  $20/17\%$  و  $55/20\%$  است. نتایج بدست آمده بیانگر موثر بودن رویکردهای پیشنهادی در حل چالش‌های سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ است.

**کلید واژه:** شرح‌گذاری خودکار تصاویر، مجموعه تصاویر در مقیاس بزرگ، انتخاب نمونه‌های اولیه، مقیاس‌پذیری و زمان شرح‌گذاری

## فهرست مطالب

### صفحه

### عنوان

.....	فهرست جدول‌ها
.....	فهرست شکل‌ها
.....	فهرست الگوریتم‌ها
.....	فهرست علایم، نشانه‌ها و مخفف‌ها
۱.....	فصل ۱: مقدمه
۲.....	۱-۱ - پیشگفتار
۲.....	۱-۲ - شرح‌گذاری خودکار تصاویر
۴.....	۱-۳ - مولفه‌های اصلی سیستم‌های شرح‌گذاری خودکار تصاویر
۶.....	۱-۴ - چالش‌های اصلی در شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ
۹.....	۱-۵ - نوآوری‌ها
۱۱.....	۱-۶ - ساختار و محتوای پایان‌نامه
۱۳.....	فصل ۲: مروری بر روش‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ
۱۴.....	۲-۱ - پیشگفتار
۱۴.....	۲-۲ - پیش‌پردازش داده‌ها
۱۷.....	۲-۳ - مدل‌های یادگیری مقیاس‌پذیر
۲۰.....	۲-۴ - توزیع‌پذیری محاسبات
۲۱.....	۲-۵ - جمع‌بندی
۲۵.....	فصل ۳: مروری بر روش‌های انتخاب نمونه‌های اولیه
۲۶.....	۳-۱ - پیشگفتار
۲۶.....	۳-۲ - انتخاب نمونه
۲۷.....	۳-۳ - انتخاب نمونه‌های اولیه

۴-۳- روش تنکسازی مبتنی بر نزدیکترین همسایه (CNN) .....	۳۰
۴-۳- روش اصلاح سازی مبتنی بر نزدیکترین همسایه (ENN) .....	۳۱
۴-۳- جمع بندی .....	۳۲
<b>فصل ۴: رویکردهای پیشنهادی</b> .....	<b>۳۳</b>
۴-۱- پیشگفتار .....	۳۴
۴-۲- انتخاب نمونه های اولیه .....	۳۴
۴-۲-۱- رویکرد پیشنهادی اول در انتخاب نمونه های اولیه .....	۳۴
۴-۲-۱-۱- نمونه های پرت طبقه ای .....	۳۵
۴-۲-۱-۲- رویکرد پیشنهادی COCNN .....	۳۶
۴-۲-۲- رویکرد پیشنهادی دوم در انتخاب نمونه های اولیه .....	۳۸
۴-۲-۲-۱- ENN چندبر چسبه .....	۳۸
۴-۲-۲-۲- عدم شباهت چندبر چسبی .....	۴۰
۴-۲-۲-۳- پیش بینی بر چسب از نزدیکترین همسایه ها .....	۴۱
<b>روش پیش بینی حریصانه</b> .....	<b>۴۳</b>
۴-۲-۴- روش پیش بینی مبتنی بر الگوریتم KELM .....	۴۴
۴-۲-۴-۱- MLENN-GKELM - روش پیشنهادی .....	۴۷
۴-۲-۴-۲- جمع بندی .....	۵۱
۴-۲-۴-۳- شرح گذاری .....	۵۱
۴-۲-۴-۴- شرح گذاری مبتنی بر K- نزدیکترین همسایه وزن دار .....	۵۲
۴-۲-۴-۵- شرح گذاری مبتنی بر الگوریتم KELM .....	۵۴
۴-۲-۴-۶- آستانه گذاری در طبقه بندی چندبر چسبه .....	۵۵
۴-۲-۴-۷- جمع بندی .....	۵۶
۴-۴- جمع بندی رویکردهای ارائه شده .....	۵۶
<b>فصل ۵: نتایج تجربی</b> .....	<b>۵۹</b>

۶۰	۱-۵ - پیشگفتار
۶۰	۲-۵ - مجموعه تصاویر مورد استفاده
۶۱	۳-۵ - توصیف تصاویر
۶۲	۴-۵ - معیارهای ارزیابی
۶۲	۴-۵ - ارزیابی نتایج شرح‌گذاری
۶۳	۴-۵ - ارزیابی نتایج انتخاب نمونه‌های اولیه
۶۴	۵-۵ - تنظیم اولیه آزمایش‌ها
۶۵	۵-۵ - تنظیم‌های انتخاب نمونه‌های اولیه
۶۹	۵-۵ - تنظیم‌های شرح‌گذاری
۷۱	۶-۵ - ارائه، تحلیل و ارزیابی نتایج تجربی
۷۲	۱-۶-۵ - نتایج NUS-WIDE-Object
۷۷	۲-۶-۵ - نتایج NUS-WIDE-Scene
۷۹	۳-۶-۵ - نتایج NUS-WIDE-Lite
۸۲	۴-۶-۵ - نتایج NUS-WIDE
۸۵	۵-۵ - تجزیه و تحلیل مقیاس‌پذیری
۸۶	۶-۵ - جمع‌بندی
۸۷	<b>فصل ۶: نتیجه‌گیری و پژوهش‌های آینده</b>
۸۸	۶-۱ - مروری بر رویکردهای پیشنهادی
۹۰	۶-۲ - مروری بر نوآوری‌ها
۹۱	۶-۳ - پژوهش‌های پیشنهادی
۹۴	واژه‌نامه فارسی به انگلیسی
۹۸	واژه‌نامه انگلیسی به فارسی
۱۰۲	مراجع

## فهرست جداول

عنوان	صفحه
جدول ۱-۲) خلاصه پژوهش‌های بررسی شده در حوزه شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ.	۲۲
جدول ۱-۴) روابط عدم شباهت دودویی متعدد.	۴۲
جدول ۱-۵) پارامترهای بهینه تنظیم شده انتخاب نمونه‌های اولیه و شرح‌گذاری.	۷۱
جدول ۲-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Object	۷۳
جدول ۳-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Scene	۷۷
جدول ۴-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Lite	۸۰
جدول ۵-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE	۸۳

## فهرست شکل‌ها

عنوان	صفحه
شکل ۱-۱: خروجی نهایی سیستم‌های شرح‌گذاری خودکار تصاویر.	۳
شکل ۱-۲: شمای کلی یک سیستم شرح‌گذاری خودکار تصویر.	۴
شکل ۱-۳: طبقه‌بندی روش‌های PS مرور شده در این پایان‌نامه.	۳۰
شکل ۱-۴: تفاوت مابین نمونه پرت و پرت طبقه‌ای.	۳۵
شکل ۲-۴: مثالی از عدم شباهت دو بردار برچسب دودویی.	۴۰
شکل ۳-۴: شمای کلی گام‌های روش MLENN-GKELM ارائه شده.	۴۸
شکل ۴-۱: اثربخشی بکارگیری روابط عدم شباهت دودویی مختلف در روش MLENN-GKELM	۶۷
شکل ۴-۲: اثربخشی بکارگیری توابع هسته مختلف در روش MLENN-GKELM	۶۸
شکل ۴-۳: اثربخشی بکارگیری رویکردهای مختلف وزن‌دهی در روش شرح‌گذاری KNN وزن‌دار.	۷۰
شکل ۴-۴: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE-Object	۷۶
شکل ۴-۵: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE-Scene	۷۹
شکل ۴-۶: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE-Lite	۸۲
شکل ۴-۷: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE	۸۴
شکل ۴-۸: نمودار زمان شرح‌گذاری هر تصویر بدون برچسب با تغییر تعداد تصاویر مجموعه آموزش.	۸۵

## فهرست الگوریتم‌ها

عنوان

صفحه

الگوریتم ۱-۴: شبه کد روش COCNN ارائه شده	۳۷
الگوریتم ۲-۴: شبه کد روش MLENN-GKELM ارائه شده	۵۱

---

## فهرست علایم، نشانه‌ها و مخفف‌ها

عنوان	علامت
Text-Based Image Retrieval	TBIR
Content-Based Image Retrieval	CBIR
Automatic Image Annotation	AIA
Instance Selection	IS
Prototype Selection	PS
Weighted K-Nearest Neighbor	WKNN
Condensed Nearest Neighbor	CNN
Multi-Label Edited Nearest Neighbor	MLENN
Edited Nearest Neighbor	ENN
Kernel Extreme Learning Machine	KELM
Support Vector Machine	SVM
Maximum Mean Discrepancy	MMD
Training Set Selection	TSS
Class Outlier based Condensed Nearest Neighbor	COCNN
Back Propagation	BP
Karush-Kuhn-Tucker	KKT
Multi Label Edited Nearest Neighbor based on Greedy and KELM prediction	MLENN-GKELM
Bag of Visual Words	BoW
Mean Average Precision	MAP

# فصل ۱: مقدمه

## ۱-۱- پیشگفتار

امروزه با رشد روزافزون استفاده از وسایل تصویربرداری رقمی<sup>۱</sup>، حجم این تصاویر روزبه روز در حال افزایش است. از طرفی با توسعه و رشد استفاده از بستر اینترنت و شبکه‌های اجتماعی، تعداد این تصاویر درون وب نیز به گونه‌ای قابل توجه افزایش یافته است. از این‌رو نیاز است تا این تصاویر به گونه‌ای مؤثر و کارا به وسیله موتورهای جستجو ساماندهی و بازیابی شوند. برای دستیابی به این هدف، سیستم‌های بازیابی تصویر متعددی ارائه شده‌اند. این حوزه پژوهشی از دهه ۱۹۷۰ تاکنون مورد توجه پژوهشگران بوده و از این‌رو در حال حاضر نیز مورد توجه پژوهشگران است [۱]. بازیابی تصاویر به دو شیوه کلی انجام می‌شود: (۱) بازیابی تصاویر مبتنی بر متن<sup>۲</sup> (TBIR) و (۲) بازیابی تصاویر مبتنی بر محتوا<sup>۳</sup> (CBIR). در CBIR ابتدا یک تصویر به عنوان پرس‌وجو<sup>۴</sup> از طرف کاربر دریافت می‌شود. سپس با توجه به ویژگی‌های دیداری<sup>۵</sup> به دست آمده از تصاویر مجموعه‌داده، ویژگی‌های دیداری تصویر پرس‌وجو استخراج می‌شود. در گام بعد شباهت بین تصویر پرس‌وجو و تصاویر مجموعه‌داده بر اساس ویژگی‌های دیداری استخراج شده، محاسبه و رتبه‌بندی نتایج بازیابی انجام می‌شود. ویژگی‌هایی مانند رنگ<sup>۶</sup>، بافت<sup>۷</sup> و شکل<sup>۸</sup> و یا ترکیب آن‌ها به عنوان ویژگی‌های دیداری سطح پایین در CBIR استفاده می‌شود. در حالی که TBIR بر اساس متون و کلیدواژه‌های نوشته شده برای هر تصویر عمل می‌کند. در این روش با آمدن هر پرس‌وجو از جانب کاربر، تصاویری که پرس‌وجوی موردنظر برای آن‌ها توصیف شده است بازیابی می‌شوند [۱].

## ۱-۲- شرح‌گذاری خودکار تصاویر

در بازیابی تصاویر بر اساس متن نیاز است تا تصاویر درون مجموعه‌داده<sup>۹</sup> شرح‌گذاری<sup>۱۰</sup> شوند. شرح‌گذاری تصاویر<sup>۱۱</sup> می‌تواند به دو صورت دستی و خودکار انجام شود. در شرح‌گذاری

<sup>1</sup> Digital

<sup>2</sup> Text-Based Image Retrieval (TBIR)

<sup>3</sup> Content-Based Image Retrieval (CBIR)

<sup>4</sup> Query

<sup>5</sup> Visual

<sup>6</sup> Color

<sup>7</sup> Texture

<sup>8</sup> Shape

<sup>9</sup> Data Set

<sup>10</sup> Annotate

<sup>11</sup> Image Annotation

دستی، تصاویر مجموعه‌داده به وسیله افراد خبره شرح‌گذاری می‌شود. از آنجایی که تعداد تصاویر درون وب و مجموعه‌داده‌های موتورهای جستجو بسیار زیاد است، انجام این کار عملاً ممکن نیست.  
از این‌رو روش‌های شرح‌گذاری خودکار تصاویر<sup>۱</sup> معرفی شده‌اند [۲].

فرآیند شرح‌گذاری خودکار تصاویر را می‌توان به صورت زیر تعریف کرد: "شرح‌گذاری خودکار تصاویر فرآیندی است که یک سیستم کامپیوتری به صورت خودکار فراداده‌هایی<sup>۲</sup> به شکل توصیف‌های متنی یا کلیدواژه‌ها<sup>۳</sup> را به تصاویر رقمی اختصاص می‌دهد. این فرآیند یکی از کاربردهای بینایی ماشین در سیستم‌های بازیابی تصاویر است و به منظور ساماندهی و مکان‌یابی تصاویر موجود در مجموعه‌ها استفاده می‌شود" [۳]. شکل ۱-۱ خروجی نهایی سیستم‌های شرح‌گذاری خودکار تصاویر را نشان می‌دهد. همان‌طور که مشاهده می‌شود، در انتهای، محتوای شرح‌گذاری خودکار تصاویر با نشان می‌دهد. همان‌طور که مشاهده می‌شود، در انتهای، محتوای مفهومی هر تصویر با چند کلمه کلیدی توصیف شده است.



شکل ۱-۱: خروجی نهایی سیستم‌های شرح‌گذاری خودکار تصاویر.

در سیستم‌های بازیابی تصاویر، شرح‌گذاری خودکار یک تصویر، به عنوان یک مرحله اولیه و پیش‌پردازشی در نظر گرفته می‌شود. هدف از این کار توصیف یک تصویر با کلمه‌های کلیدی و توصیف‌های متنی بر اساس محتوای دیداری هر تصویر است. در این سیستم‌ها شرح‌گذاری تصویر یک نقطه عطف به حساب می‌آید. زیرا روش‌های شرح‌گذاری خودکار تصاویر با استفاده از محتوای دیداری تصویر فرآیند شرح‌گذاری را انجام می‌دهند و از این طریق با بهره‌گیری از مزایای بازیابی تصاویر مبتنی بر محتوا، جستجوی یک تصویر به فرآیند تطابق متن<sup>۴</sup> تبدیل می‌شود. این کار سرعت بازیابی تصویر را به مقدار قابل توجهی افزایش می‌دهد [۴].

<sup>۱</sup> Automatic Image Annotation(AIA)

<sup>۲</sup> Meta Data

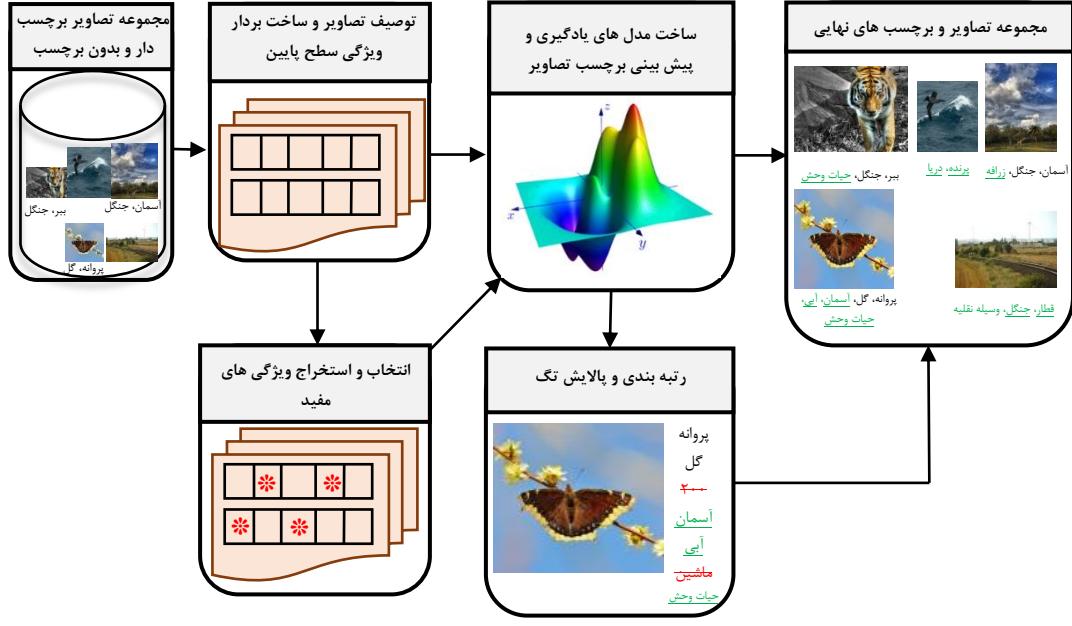
<sup>۳</sup> Keywords

<sup>۴</sup> Text Matching

موتورهای جستجوی فعلی از متون اطراف تصاویر برای بازیابی تصاویر مبتنی بر متن استفاده می‌کنند. استفاده از متن اطراف تصاویر معایبی را به همراه دارد. اول اینکه محتوای درون و ب در حال تغییر است و منبع قابل اعتمادی برای استفاده از آن در تطبیق متن پرس‌وجوی کاربران نیست. دیگر اینکه این اطلاعات تضمینی برای توصیف صحیح محتوای تصویر نمی‌دهند. پس شرح‌گذاری خودکار تصاویر نتایج جستجوی کاربران در موتورهای جستجو را دقیق‌تر و قابل اعتمادتر خواهد کرد [۲]. به طور کلی می‌توان گفت که هدف کلی شرح‌گذاری خودکار تصاویر کمک به بهبود دقت<sup>۱</sup> سیستم‌های بازیابی تصاویر مبتنی بر متن است.

### ۳-۱- مولفه‌های اصلی سیستم‌های شرح‌گذاری خودکار تصاویر

شکل ۲-۱ شماتیک کلی یک سیستم شرح‌گذاری خودکار تصویر را نشان می‌دهد. بطور کلی این سیستم‌ها شامل ۶ قسمت اصلی است که در ادامه به توصیف آن‌ها می‌پردازیم.



شکل ۲-۱: شماتیک کلی یک سیستم شرح‌گذاری خودکار تصویر.

<sup>۱</sup> Precision

**الف) مجموعه تصاویر برچسبدار و بدون برچسب:** این قسمت یک مخزن<sup>۱</sup> شامل تعدادی از تصاویر برچسبدار و بدون برچسب می‌باشد. با افزایش تعداد تصاویر درون وب و پایگاه‌های اشتراک‌گذاری تصاویر، این مخزن نیز به شدت رشد می‌کند. همچنین برچسب‌های جمع‌آوری شده برای تصاویر، زیاد، متعدد، ناکامل و غیر قابل اعتماد هستند. افزایش تعداد تصاویر و کلمه‌های کلیدی و ناکامل بودن آن‌ها، سیستم‌های شرح‌گذاری تصویر را با چالش‌های جدی مواجه می‌کند. عمدۀ این چالش‌ها، مقیاس‌پذیری سیستم، زمان و دقت شرح‌گذاری هستند. در سال‌های اخیر برای حل مسائل پیش‌روی سیستم‌های شرح‌گذاری تصویر با تعداد تصاویر و کلمه‌های شرح‌گذاری زیاد، پژوهش‌هایی صورت گرفته است. تمرکز اصلی این پایان‌نامه نیز بررسی و حل چالش‌های سیستم‌های شرح‌گذاری خودکار تصاویر با وجود تعداد زیاد تصویر است.

**ب) توصیف تصاویر و ساخت بردار ویژگی سطح پایین:** در این قسمت از سیستم، به ازای هر تصویر یک بردار ویژگی سطح پایین ساخته می‌شود. ویژگی‌های دیداری سطح پایین تصاویر را می‌توان به دو دسته سراسری<sup>۲</sup> و محلی<sup>۳</sup> تقسیم‌بندی کرد [۵]. بر اساس رویکرد روش یادگیری، می‌توان هر کدام از ویژگی‌های سراسری، محلی و یا ترکیبی از آن‌ها را مورد استفاده قرار داد.

**ج) انتخاب و استخراج ویژگی‌های مفید:** این قسمت و ساخت مدل یادگیری از مهم‌ترین قسمت‌های تشکیل‌دهنده یک سیستم شرح‌گذاری خودکار تصاویر هستند. اهمیت انتخاب ویژگی از آن جهت است که پیچیدگی بیشتر الگوریتم‌های یادگیری ماشین وابسته به تعداد ویژگی‌های تصاویر و همچنین تعداد تصاویر موجود در مجموعه‌ها است. بنابراین برای کاهش حافظه مصرفی و محاسبه‌ها، کاهش ابعاد بردار ویژگی از اهمیت بالایی برخوردار است. از طرفی برخی از ویژگی‌ها موجود قدرت لازم و کافی برای تفکیک و تمایز داده‌ها را ندارند. در مقابل تعدادی از ویژگی‌ها قدرت فراوانی در تفکیک‌پذیری داده‌ها دارند [۶]. از این‌رو در بیشتر کاربردها، انتخاب و استخراج ویژگی از اهمیت ویژه‌ای برخوردار است.

<sup>1</sup> Repository

<sup>2</sup> Global

<sup>3</sup> Local

**د) ساخت مدل‌های یادگیری و انجام عملیات شرح‌گذاری:** در این قسمت، از ویژگی‌های به دست آمده از قسمت قبلی، مدل‌های یادگیری ساخته شده و بر اساس رویکرد مدل یادگیری عملیات پیش‌بینی برچسب و شرح‌گذاری انجام می‌شود. رویکردهای یادگیری برای شرح‌گذاری تصاویر می‌تواند پارامتری<sup>۱</sup> و یا غیرپارامتری<sup>۲</sup> باشند. از طرفی این رویکردها می‌توانند روش‌های یادگیری بانظارتی<sup>۳</sup> و یا نیمه‌نظرارتی باشند [۶]. در انتهای این قسمت به ازای هر تصویر با توجه به تعداد کلمه‌های شرح‌گذاری برای هر کاربرد خاص، چندین کلمه تولید می‌شود. این کلمه‌ها توصیف‌کننده محتوای دیداری تصویر هستند.

**۵) رتبه‌بندی و پالایش تگ‌ها:** این مرحله در برخی از سیستم‌ها به عنوان یک مرحله پس‌پردازشی<sup>۴</sup> استفاده می‌شود. در این مرحله برچسب‌های تولیدشده توسط مدل‌های یادگیری مورد تجزیه و تحلیل قرار گرفته و پالایش مجدد می‌شوند. انجام این کار در این سیستم‌ها برای افزایش دقت و عملکرد<sup>۵</sup> نتایج تولیدشده توسط سیستم است [۷]. در برخی از کاربردها فرآیند رتبه‌بندی و پالایش برچسب‌ها به صورت یک سیستم جداگانه است. در این موارد، در مجموعه تصاویر اولیه همه تصاویر برچسبدار بوده و تصویر بدون برچسب موجود نیست. در این مجموعه تصاویر برچسب‌ها معمولاً غیردقیق و ناکامل هستند. روش‌های پالایش تک سعی در بهبود تگ‌های موجود برای تصاویر دارند [۸].

## ۱-۴- چالش‌های اصلی در شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ

همان‌طور که پیش‌تر نیز گفته شد، تمرکز اصلی این پایان‌نامه شرح‌گذاری خودکار تصاویر با وجود تعداد زیاد تصاویر و کلمه‌های شرح‌گذاری درون مجموعه است. از این‌رو در این قسمت چالش‌های اصلی رودرروی سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ<sup>۶</sup>

<sup>1</sup> Parametric

<sup>2</sup> Non-parametric

<sup>3</sup> Supervised

<sup>4</sup> Post processing

<sup>5</sup> Performance

<sup>6</sup> Large Scale

مورد بررسی قرار می‌گیرند. برای طراحی یک سیستم شرح‌گذاری خودکار تصویر در مقیاس بزرگ باید معیارهای زیر را در نظر گرفت:

- ۱) اولین معیار مهم وسعت و مقیاس<sup>۱</sup> سیستم است. با وجود پیشرفتهای و تکنیکهای تصویربرداری رقمی، تصاویر مختلفی در حوزه‌های متفاوت تولید می‌شوند. با در نظر گرفتن برچسب‌های مفهومی تصاویر به عنوان طبقه‌ها، در مرحله ۴ از شکل ۲-۱، مجموعه‌ای از طبقه‌بند<sup>۲</sup>ها برای شرح‌گذاری تصاویر آموزش<sup>۳</sup> داده می‌شوند. پژوهش‌های متعددی در مجموعه‌داده‌های با مقیاس متوسط و کوچک مانند Corel [۹] انجام شده است. اما مدل‌های ارائه شده در این پژوهش‌ها در شرح‌گذاری تعداد زیاد تصاویر با حوزه‌های وسیع و متفاوت مفهومی با محدودیت‌های بسیاری مواجه می‌شوند. یک سیستم شرح‌گذاری مناسب باید امکان شرح‌گذاری هر تعداد تصویر با حوزه‌های گسترده و متفاوت را داشته باشد. به عبارت دیگر این سیستم باید قابلیت مقیاس‌پذیری<sup>۴</sup> و گسترش یافتن دامنه تصاویر را داشته باشد [۱۰].
- ۲) دومین معیار مهم در طراحی یک سیستم شرح‌گذاری، زمان صرف شده برای شرح‌گذاری هر تصویر بدون برچسب است. مدت زمان انتظار کاربر در فرآیند شرح‌گذاری و پاسخ به پرس‌وجو از پارامترهای مهم یک سیستم برخط<sup>۵</sup> است. از این‌رو زمان شرح‌گذاری نیز یک معیار کلیدی و مهم در سیستم‌های شرح‌گذاری تصاویر است. برخی از روش‌های شرح‌گذاری موجود، این فرآیند را با یافتن تصاویر مشابه یک تصویر بدون برچسب و شرح‌گذاری تصویر بدون برچسب بر اساس برچسب‌های تصاویر مشابه انجام می‌دهند. با این حال مجموعه تصاویر در مقیاس بزرگ شامل هزاران و میلیون‌ها تصویر است. در این روش‌ها با افزایش مقیاس مجموعه داده‌ها، مشخصه بلادرنگ<sup>۶</sup> بودن سیستم تضمین نمی‌شود. بنابراین یکی از نیازهای اساسی یک سیستم شرح‌گذاری بلادرنگ بودن آن است [۱۰].

<sup>1</sup> Scale

<sup>2</sup> Classifier

<sup>3</sup> Train

<sup>4</sup> Scalability

<sup>5</sup> Online

<sup>6</sup> Real time

۳) نهایتاً سومین معیار دقت شرح‌گذاری است [۱۰]. همان‌طور که بیان شد، در هر کدام از مراحل سیستم راهکاری برای بهبود دقت وجود دارد. برای مثال در مرحله توصیف تصویر می‌توان از روش‌های استخراج ویژگی که ویژگی‌های تفکیک کننده را استخراج می‌کنند، استفاده کرد. در مرحله یادگیری می‌توان از مدل‌های یادگیری قدرتمند استفاده کرد. یا در نهایت پس از شرح‌گذاری و تولید برچسب نهایی از پالایش تک برای افزایش دقت بهره برد.

پس به طور کلی برای طراحی یک سیستم شرح‌گذاری تصویر با مجموعه تصاویر مقیاس بزرگ، باید سه چالش و مشخصه مهم مقیاس‌پذیری، زمان شرح‌گذاری و دقت شرح‌گذاری در نظر گرفته شوند. به منظور حل این چالش‌ها، راه حل‌های مختلفی ارائه شده است. پژوهش‌هایی در جهت بهبود دقت شرح‌گذاری و همچنین حل چالش زمان شرح‌گذاری و مقیاس‌پذیری سیستم از جمله این راه حل‌ها می‌باشند. یکی از رویکردهایی که تابحال در حل چالش دقت و زمان شرح‌گذاری در این سیستم‌ها پرداخته نشده است، کاهش حجم مجموعه آموزش و حذف نمونه‌های آموزش بی‌تأثیر در فرآیند طبقه‌بندی است. به این فرآیند "انتخاب نمونه"<sup>۱</sup> (IS) می‌گویند. به عبارت دقیق‌تر IS، فرآیند یافتن نماینده‌های هر طبقه<sup>۲</sup> و حذف نمونه‌های آموزش بی‌تأثیر در فرآیند طبقه‌بندی است. پس از شناسایی و انتخاب نمونه‌های مناسب، زیرمجموعه‌ای از مجموعه آموزش و با حجمی کمتر به دست می‌آید. بدین ترتیب زیرمجموعه انتخاب شده به عنوان مجموعه آموزش جدید روش‌های یادگیری مورد استفاده قرار می‌گیرد. بکارگیری این ایده منجر به کاهش حجم مجموعه آموزش و حذف نمونه‌های دارای اختلال<sup>۳</sup> و بی‌تأثیر در فرآیند طبقه‌بندی می‌شود. نتیجه این کار کاهش زمان شرح‌گذاری و افزایش دقت شرح‌گذاری است.

در این پایان‌نامه سه چالش اصلی پیش روی سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ مورد بررسی قرار گرفته و راهکارهایی در جهت حل هر یک پیشنهاد شده است. چالش‌های مقیاس‌پذیری و زمان شرح‌گذاری با بکارگیری روش‌های انتخاب نمونه‌های

<sup>1</sup> Instance Selection (IS)

<sup>2</sup> Class

<sup>3</sup>Noisy

اولیه<sup>۱</sup> (PS) و ارائه نسخه‌های چندبرچسبه و بهبود یافته از این روش مورد بررسی قرار گرفته است. از طرفی دقت شرح‌گذاری نیز با بکارگیری روش‌های پیش‌بینی برچسب ماشین یادگیری سرعی<sup>۲</sup> (ELM) مبتنی بر هسته‌های غیر خطی و K-نزدیکترین همسایه وزن‌دار<sup>۳</sup> (WKNN) بهبود یافته است.

## ۱-۵- نوآوری‌ها

در این پایان‌نامه نوآوری‌های زیر ارائه شده است.

(۱) بکارگیری و بهبود روش انتخاب نمونه‌های اولیه CNN در کاربرد شرح‌گذاری خودکار تصاویر: روش‌های PS در کاربردهای متعددی استفاده شده است. فرضیه اصلی در این روش‌ها تنک‌سازی نمونه‌ها مبتنی بر عدم‌شباهت هر نمونه با همسایگی‌هایش می‌باشد. روش CNN<sup>۴</sup> یکی از روش‌های پایه در این حوزه می‌باشد. با توجه به اینکه از ایده انتخاب نمونه‌های اولیه برای حل چالش مقیاس‌پذیری روش‌های شرح‌گذاری خودکار تصاویر استفاده نشده است، این ایده نوآوری اولیه این پایان‌نامه می‌باشد. نتایج حاصل از اثربخشی این فرضیه در [۱۱] ارائه شده است. همچنین در [۱۱] برای افزایش نرخ کاهش نمونه‌ها و یافتن زیرمجموعه کاهش یافته کمینه از ایده حذف نمونه‌های پرت طبقه‌ای<sup>۵</sup> استفاده شده است.

(۲) بهبود روش انتخاب نمونه‌های اولیه چندبرچسبه MLENN با اثربخشی روابط عدم‌شباهت مابین بردار برچسب نمونه‌های کاندیدای حذف: هرچند CNN روشی پایه‌ای در انتخاب نمونه‌های اولیه است، ولی این روش منطبق بر کاربردهای چندبرچسبه<sup>۶</sup> نمی‌باشد. در کاربردهای چندبرچسبه مانند شرح‌گذاری خودکار تصاویر، هر نمونه می‌تواند به بیش از یک کلاس تعلق داشته باشد. از طرفی CNN روشی تکرار شونده و کند بوده و در کاربردهای مقیاس بزرگ با

<sup>1</sup> Prototype Selection (PS)

<sup>2</sup> Extreme Learning Machine (ELM)

<sup>3</sup> Weighted K-Nearest Neighbor (WKNN)

<sup>4</sup> Condensed Nearest Neighbor (CNN)

<sup>5</sup> Class outlier

<sup>6</sup> Multi-label

محدودیت زمان اجرا مواجه می‌شود. روش MLENN<sup>۱</sup> که نسخه توسعه یافتهENN<sup>۲</sup> است، تنها روشی است که برای کاربردهای چندبرچسبه ارائه شده است. همچنین پیچیدگی محاسباتی این روش خطی بوده و دارای مرحله اجرای ثابت است. در این روش برای محاسبه عدم شباهت<sup>۳</sup> بین بردارهای برچسب از فاصله همینگ<sup>۴</sup> استفاده می‌شود. دومین فرضیه اساسی در این پایان‌نامه، اثربخشی رویکرد محاسبه عدم شباهت بین بردار برچسب نمونه‌ها در افزایش نرخ کاهش در راستای مقیاس‌پذیری می‌باشد. همچنین در این فرضیه با بکارگیری ایده روابط عدم شباهت تنظیم شده<sup>۵</sup> [۱۲]، روابط عدم شباهت موجود تغییر یافته و در فرضیه ارائه شده مورد استفاده قرار گرفته است.

نتایج حاصل از موثربودن این فرضیه در [۱۳] آورده شده که در مرحله بازنگری می‌باشد.

۳) بهبود روش انتخاب نمونه اولیه چندبرچسبه MLENN برپایه الگوریتمی حریصانه در پیش‌بینی بردار برچسب نمونه‌های کاندیدای حذف: در MLENN برای تصمیم‌گیری حذف هر نمونه، از رای‌گیری حداکثری<sup>۶</sup> برای پیش‌بینی بردار برچسب همسایگی‌های آن استفاده می‌شود. سومین فرضیه این پایان‌نامه اثربخشی پیش‌بینی بردار برچسب از همسایگی‌ها در بهبود عملکرد انتخاب نمونه‌های اولیه می‌باشد. در این راستا الگوریتمی ارائه شده است که پیش‌بینی برچسب هر نمونه از همسایگی‌هایش را با رویکردی حریصانه انجام می‌دهد. نتایج حاصل از اثربخشی این فرضیه در [۱۴] آورده است که در مرحله داوری می‌باشد.

۴) بهبود دقت شرح‌گذاری تصاویر بدون برچسب برپایه ماشین یادگیری سریع مبتنی بر هسته‌های غیرخطی: هرچند طی نوآوری‌های ارائه شده در مراحل قبلی، مقیاس‌پذیری افزایش یافت ولی در دقت شرح‌گذاری بهبودی حاصل نشد. با توجه به موثر بودن روش ماشین یادگیری سریع مبتنی بر هسته غیر خطی<sup>۷</sup> (KELM) در کاربردهای متعدد، فرضیه اثربخش بودن آن در

<sup>1</sup> Multi-Label Edited Nearest Neighbor (MLENN)

<sup>2</sup> Edited Nearest Neighbor (ENN)

<sup>3</sup> Distance

<sup>4</sup> Hamming

<sup>5</sup> Adjusted

<sup>6</sup> Majority Voting

<sup>7</sup> Kernel Extreme Learning Machine (KELM)

شرحگذاری خودکار تصاویر در مقیاس بزرگ نیز در این پژوهش آزمون گردید. نتایج حاصل از اثربخشی این فرضیه در [۱۵] ارائه گردیده است.

(۵) ارائه روشی نو در انتخاب نمونه‌های اولیه در کابردهای چندبرچسبه برپایه ماشین یادگیری سریع مبتنی بر هسته غیرخطی: با توجه موثر بودن پیش‌بینی برچسب در بهبود نرخ کاهش نمونه‌ها [۱۴] و همچنین موثر بودن KELM در افزایش دقت شرحگذاری [۱۵]، فرضیه استفاده از KELM در پیش‌بینی با شکاف معنایی کمتر در بردار برچسب‌ها نیز مطرح گردید. نتایج حاصل از اثربخشی این فرضیه در [۱۶] آورده شده است که در مرحله داوری می‌باشد.

در نهایت عملکرد مطلوب نوآوری‌های ارائه شده بر روی مجموعه تصاویر مقیاس بزرگ متعدد مورد آزمایش قرار گرفته و با سایر پژوهش‌های صورت گرفته در حوزه شرحگذاری خودکار تصاویر در مقیاس بزرگ و انتخاب نمونه‌های اولیه مورد مقایسه قرار گرفته است.

## ۱-۶- ساختار و محتوای پایان‌نامه

از آنجایی که تمرکز اصلی این پایان‌نامه بررسی و حل چالش‌های شرحگذاری خودکار مجموعه تصاویر در مقیاس بزرگ می‌باشد، در اولین فصل پس از مقدمه، مروری بر پژوهش‌های صورت گرفته در حوزه شرحگذاری خودکار مجموعه تصاویر در مقیاس بزرگ خواهیم داشت. در این پایان‌نامه به منظور حل چالش‌های مطرح شده در حوزه شرحگذاری از روش‌های PS استفاده شده است. از این‌رو در فصل سوم، مروری بر روش‌های PS صورت گرفته و طبقه‌بندی این روش‌ها ارائه خواهد شد. فصل چهارم به رویکردهای ارائه شده در این پایان‌نامه در جهت حل چالش‌های سیستم‌های شرحگذاری خودکار تصاویر در مقیاس بزرگ پرداخته شده و روش پیشنهادی ارائه می‌شود. فصل پنجم به چگونگی پیاده‌سازی و بیان نتایج تجربی بر روی چهار دادگان می‌پردازد. نتایج ارائه شده در این فصل نشان از صلاحیت رویکردهای ارائه شده در حل چالش‌های مطرح شده دارند. فصل ششم نیز جمع‌بندی کوتاهی از مطالب ارائه شده و پیشنهادهایی برای پژوهش‌های آینده خواهد داشت. در نهایت منابع و مأخذ و واژه‌نامه انگلیسی-فارسی و فارسی - انگلیسی پایان بخشنده این پایان‌نامه خواهند بود.

# فصل ۲:

## مرواری بر روش‌های شرح‌گذاری

### خودکار مجموعه تصاویر در

### مقیاس بزرگ

## ۱-۲- پیشگفتار

همانطور که بیان شد چالش‌های اصلی پیش‌روی سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ، مقیاس‌پذیری، زمان شرح‌گذاری و دقت شرح‌گذاری هستند. به منظور حل این چالش پژوهش‌های گستردۀ‌ای صورت گرفته است. پژوهش‌های صورت گرفته را می‌توان به سه دسته پیش‌پردازش داده‌ها، بکارگیری و ارائه مدل‌های یادگیری مقیاس‌پذیر و توزیع پذیری محاسبات تقسیم بندی نمود. بطور کلی در پژوهش‌های صورت گرفته در این حوزه، تمرکز اصلی بر روی مقیاس‌پذیری و زمان شرح‌گذاری بوده و دقت شرح‌گذاری بصورت همزمان بهبود می‌یابد. به عنوان مثال، در حوزه انتخاب و استخراج ویژگی، دقت و مقیاس‌پذیری سیستم بصورت همزمان بهبود می‌یابد. از این‌رو طبقه‌بندی روش‌های بررسی شده در این فصل بر اساس مقیاس‌پذیری و پیچیدگی سیستم بوده و در هر روش به صورت جداگانه به چگونگی بهبود دقت شرح‌گذاری اشاره خواهد شد. در ادامه مروری بر کارهای پژوهشی که در هر حوزه انجام شده است آورده می‌شود.

## ۲-۲- پیش‌پردازش داده‌ها

در برخی از پژوهش‌های صورت گرفته چالش‌های مقیاس‌پذیری و دقت شرح‌گذاری با انجام فرآیند پیش‌پردازش بر روی داده‌ها مورد بررسی قرار گرفته است. در این پژوهش‌ها در ابتدا داده‌های مقیاس بزرگ مورد پیش‌پردازش قرار گرفته و سپس با الگوریتم‌های یادگیری ماشین پایه، عملیات شرح‌گذاری تصاویر صورت می‌گیرد. از معروفترین فرآیندهای پیش‌پردازش داده‌ها انتخاب ویژگی و انتخاب نمونه است. انتخاب ویژگی فرآیند یافتن موثرترین ویژگی‌ها، از بین ویژگی‌های موجود و حذف ویژگی‌های افزونه است. از طرفی انتخاب نمونه فرآیند یافتن نمونه‌های موثر و حذف نمونه‌های بی‌تأثیر در فرآیند طبقه‌بندی است. همانطور که بیان شد ایده انتخاب نمونه‌ها در حل چالش مقیاس‌پذیری و زمان شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ تا حال مورد بررسی قرار نگرفته است. از این‌رو ایده اصلی این پایان‌نامه پیش‌پردازش داده‌ها با انتخاب نمونه‌های موثر در فرآیند طبقه‌بندی در شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ می‌باشد. از طرفی در پژوهش‌های متعدد از جمله [۱۰]-[۱۷]-[۲۳] چالش زمان و دقت

شرح‌گذاری با استفاده از پیش‌پردازش نمونه‌ها حل شده است. از این‌رو در ادامه این بخش این پژوهش‌ها مورد بررسی قرار می‌گیرند.

در [۱۰] برای انتخاب ویژگی‌های مفید از الگوریتم ژنتیک<sup>۱</sup> بهره برده شده است. الگوریتم ژنتیک یک روش برنامه‌نویسی است که از اصل تکامل ژنتیکی برای حل مسئله استفاده می‌کند. ورودی الگوریتم مسئله، و خروجی، راه حل‌های کدگذاری شده طبق یک الگوی<sup>۲</sup> مشخص می‌باشد. در این مقاله از الگوریتم ژنتیک مبتنی بر کروموزوم‌های کد شده دودویی<sup>۳</sup> برای انتخاب ویژگی‌های مفید استفاده شده است. در هر مرحله از الگوریتم بر اساس تابع برازش، که دقت به دست آمده توسط یک الگوریتم طبقه‌بندی است، کروموزوم‌ها بروز شده تا درنهایت به یک برداری که نشان‌دهنده وزن هر ویژگی در فضای ویژگی مربوطه است، برسد. درنهایت نیز بر اساس بردار وزنی به دست آمده، زیرمجموعه ویژگی دارای وزن‌های بالاتر انتخاب می‌شود.

ما<sup>۴</sup> و همکارانش در [۱۸] استدلال کرده‌اند که بیشتر روش‌های استخراج ویژگی مطرح شده مانند رتبه فیشر<sup>۵</sup>، ویژگی‌ها را یکی‌یکی و بر اساس اهمیتشان مرتب کرده و انتخاب می‌کنند. این روش‌ها ارتباط موجود مابین ویژگی‌ها را در نظر نمی‌گیرند. از این‌رو دسته دیگری از روش‌ها بنام انتخاب ویژگی پراکنده<sup>۶</sup> مطرح شده‌اند.

هدف از انتخاب ویژگی پراکنده استفاده از یک مدل پراکنده و تنک برای انتخاب مفیدترین ویژگی‌هایی است که قدرت بالایی در تفکیک کردن داده‌ها از هم دارند. در این روش‌ها که یکی از معروف‌ترین آن‌ها نرم ماتریسی<sup>۷</sup>  $L_{2,1}$  است، یک ماتریس  $W$  پراکنده یافت شده و ویژگی‌های متناظر با سطرهای غیر صفر این ماتریس انتخاب می‌شوند. انتخاب ویژگی مبتنی بر پراکنده‌گی روشی مؤثر و کارا است زیرا ارتباط موجود مابین ویژگی‌ها را در نظر می‌گیرد که باعث بهبود

<sup>1</sup> Genetic

<sup>2</sup> Pattern

<sup>3</sup> Bi-coded

<sup>4</sup> Ma

<sup>5</sup> Fisher score

<sup>6</sup> Sparse Feature Selection

<sup>7</sup> Matrix Norm

عملکرد الگوریتم یادگیری می‌شود. علاوه بر این هزینه محاسباتی الگوریتم‌ها را تا حد قابل توجهی کاهش می‌دهد [۱۷].

استدلال دیگر مقاله [۱۸] این است که روش‌هایی که مسئله چندبرچسبه شرح‌گذاری تصاویر را به چند مسئله تک برچسبه تبدیل می‌کنند، از ارتباط موجود مابین برچسب‌ها صرف نظر می‌کنند. ایده نوآورانه این مقاله در این است که از انتخاب ویژگی پراکنده مبتنی بر نرم ماتریسی  $L_{2,1}$  استفاده کرده است. همچنین ماتریس زیر فضای اشتراکی موجود مابین کلاس‌ها به صورت ترکیبی در فرموله‌سازی مسئله پیش‌بینی برچسب استفاده می‌شود. از طرفی به طور همزمان ماتریس پراکنده  $W$  از طریق فرموله‌سازی مربوطه یاد گرفته می‌شود. این مقاله در فرموله‌سازی مسئله از خطای حداقل مربعات استفاده کرده و با استفاده از یک الگوریتم تکرارشونده مسئله بهینه‌سازی را حل کرده است. درنهایت با انجام آزمایش‌هایی برروی دو مجموعه تصویر مقیاس بزرگ و مقایسه روش ارائه شده با دیگر روش‌های انتخاب ویژگی، صلاحیت روش ارائه شده نشان داده شده است.

در [۱۷] نیز به مسائل مربوط به روش‌های انتخاب ویژگی و در نظر نگرفتن ارتباط مابین ویژگی‌ها می‌پردازد. راهکار این مقاله نیز استفاده از انتخاب ویژگی پراکنده می‌باشد. ایده اصلی این مقاله بکارگیری نرم  $L_{2,1/2}$  به منظور انتخاب ویژگی پراکنده است. این مقاله از این روش انتخاب ویژگی برای کاربرد شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ بهره برده و یک قاب‌کاری<sup>۱</sup> جدید مطرح کرده است. در این قاب‌کاری مطرح شده برای انجام فرآیند شرح‌گذاری از روش یادگیری نیمه‌نظری مبتنی بر گراف لابلسی<sup>۲</sup> استفاده شده است.

به منظور حل چالش مقیاس‌پذیری در سیستم‌های شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ از روش هش<sup>۳</sup> کردن ویژگی‌ها و نمونه‌ها نیز بهره برده شده است [۱۹]-[۲۲]. چالش تعداد بالای برچسب‌ها در انتشار برچسب مبتنی بر گراف عمده تمرکز مقاله [۲۲] می‌باشد. بکارگیری

<sup>1</sup> Framework

<sup>2</sup> Graph Laplacian-based Semi-Supervised Learning

<sup>3</sup> Hash

ایده هش محلی<sup>۱</sup> در ساخت گرافهای مقیاس پذیر ایده اصلی این مقاله در حل چالش مقیاسپذیری است.

کاهش ابعاد فضای بردار ویژگی با استفاده از توابع هش ایده اصلی مقاله [۲۱] در مقیاس پذیر کردن روش ارائه شده می‌باشد. در روش ارائه شده مجموعه‌ای از توابع هش دودویی در تبدیل بردار ویژگی فضای اصلی به فضای کاهش یافته مورد استفاده قرار گرفته است. از این‌رو پیچیدگی‌های ذخیره‌سازی و زمان شرح‌گذاری کاهش یافته است. مقالات [۱۹] و [۲۰] نیز از ایده‌های مشابه یعنی هش کردن نمونه‌ها و ویژگی‌ها بهره برده‌اند.

در نهایت مقاله [۲۳] نیز به منظور غلبه بر محدودیت‌های مقیاس بزرگ از ایده پیش‌پردازش نمونه‌ها بهره برده است. در این مقاله پیچیدگی زمانی انتشار برچسب در روش یادگیری نیمه‌نظرارتی مبتنی بر گراف مورد بررسی قرار گرفته است. برای حل محدودیت بیان شده در این مقاله، در ابتدا برخی نمونه‌های منتخب یافت می‌شوند. سپس گراف نمونه‌های منتخب بجای گراف نمونه‌های اصلی ساخته می‌شود. در این الگوریتم، خوشه‌بندی K میانگین بر روی نمونه‌های اصلی اجرا شده و مرکز خوشه‌ها به عنوان نمونه‌های منتخب در نظر گرفته می‌شوند.

### ۳-۲- مدل‌های یادگیری مقیاس‌پذیر

در برخی از پژوهش‌های صورت گرفته محدودیت و چالش مقیاس‌پذیری با ارائه مدل‌های یادگیری مقیاس‌پذیر بهبود یافته است. [۲۴]، [۱۰] و [۲۶]-[۲۴] پژوهش‌های بررسی شده در این بخش را می‌توان به دو دسته بکارگیری الگوریتم‌های بانظارت و نیمه‌نظرارتی تقسیم‌بندی کرد. در روش‌های یادگیری بانظارت از روش‌های یادگیری ماشین، سیستم با استفاده از مجموعه داده‌های آموزش که شامل جفت‌های ورودی-خروجی هستند، یک نگاشت از ورودی به خروجی یاد می‌گیرد. این نگاشت در پیش‌بینی برای داده‌های جدید مورد استفاده قرار می‌گیرد. با ورود هر داده جدید برچسب آن توسط این مدل یاد گرفته شده پیش‌بینی می‌شود [۶].

<sup>۱</sup> Local Hashing

در [۲] برای شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ یک قاب‌کاری با رویکرد یادگیری بانظارت ارائه شده است. در این پژوهش در ابتدا الگوریتمی جهت انتخاب تصاویر آموزش مناسب و حذف تصاویر دارای مشکل مطرح شده است. سپس با استفاده از روش K-نzdیکترین همسایه وزن‌دار، تصاویر بدون برچسب شرح‌گذاری می‌شوند. در نهایت نیز با استفاده از سطوح کلمات WordNet، برچسب‌های تولید شده پالایش می‌شوند.

در [۱۰] از روش ماشین بردار پشتیبان<sup>۱</sup> (SVM) در شرح‌گذاری خودکار تصاویر در مقیاس بزرگ استفاده شده است. برای یک مسئله طبقه‌بندی دو برچسبه، SVM سعی دریافتن بهترین ابر صفحه<sup>۲</sup> در فضای ویژگی دارد که نمونه‌های دو برچسب را از هم جدا کند. پس برای یک مسئله طبقه‌بندی  $N$  برچسبه به  $\frac{N*(N-1)}{2}$  طبقه‌بند نیاز است. برای تعداد برچسب‌های زیاد که یکی از مشخصه‌های مجموعه تصاویر در مقیاس بزرگ است، تعداد طبقه بندها بسیار زیاد می‌شود. در [۱۰] این مشکل با ارائه مفهوم ناحیه همسایگی محلی حل شده است. ناحیه همسایگی محلی، K-نzdیکترین تصویر به مرکز فضای ویژگی هر برچسب است. ایده اصلی این مقاله یادگیری طبقه بند SVM با تصاویر درون هر ناحیه همسایگی محلی است. در مرحله شرح‌گذاری، ابتدا ناحیه همسایگی محلی هر تصویر بدون برچسب یافت شده و سپس از طبقه‌بندهای موجود در آن رای‌گیری و شرح‌گذاری نهایی انجام می‌شود.

الگوریتم‌های بانظارت برای ساخت مدل و انجام عملیات طبقه‌بندی نیاز به مقدار قابل توجهی داده برچسب‌دار با برچسب‌های قابل اعتماد و کامل دارند. برچسب‌گذاری نمونه‌ها نیز اغلب سخت و هزینه‌بر است. از آنجایی که جمع‌آوری داده‌های بدون برچسب ساده است، روش‌های نیمه‌نظرارتی برای بکارگیری این داده‌ها در ساخت مدل و انجام فرآیند طبقه‌بندی مطرح شده‌اند. به بیان دیگر روش‌های نیمه‌نظرارتی از تعداد کمی داده برچسب‌دار و ترکیب با داده‌های بدون برچسب برای ساخت مدل و انجام عملیات طبقه‌بندی استفاده می‌کنند [۲۷].

<sup>1</sup> Support Vector Machine (SVM)

<sup>2</sup> Hyper Plane

در شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ نیز این موضوع صادق است. به بیان دقیق‌تر، در مجموعه تصاویر با مقیاس بزرگ، تعداد کمی تصویر شرح‌گذاری شده با شرح‌های کامل و قابل اعتماد وجود دارد. تعداد زیادی تصویر بدون شرح نیز در مجموعه‌ها یافت می‌شود. برخی از پژوهش‌های این حوزه نیز از روش‌های یادگیری نیمه‌نظراتی برای شرح‌گذاری مجموعه تصاویر با مقیاس بزرگ استفاده کرده‌اند [۴]، [۲۵] و [۲۶].

در [۴] از روش یادگیری نیمه‌نظراتی مبتنی بر گراف در شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ استفاده شده است. چالش اصلی این مقاله نادیده گرفتن ارتباط مابین برچسب‌ها در روش‌های دیگر است. به عبارت دقیق‌تر روش‌های اخیر مطرح شده ، مسئله شرح‌گذاری را به یک مسئله طبقه‌بندی چند طبقه‌ای<sup>۱</sup> تبدیل می‌کنند. اما در این روش‌ها از ارتباط موجود مابین برچسب‌ها صرف‌نظر می‌شود. از این‌رو این مقاله فرض می‌کند یک ساختار مشترک مابین برچسب‌های موجود برای شرح‌گذاری وجود دارد. به عبارت دیگر این مقاله با بهره‌گیری از تصاویر بدون برچسب و اطلاعات همبستگی<sup>۲</sup> مابین برچسب‌ها، قاب‌کاری شرح‌گذاری تصویر خود را مطرح کرده است.

در [۲۵] رویکرد انتشار چند برچسبی در یادگیری نیمه‌نظراتی مبتنی بر گراف در شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ ارائه شده است. به ادعای این مقاله، پژوهش‌های قبلی انجام شده روی انتشار یک برچسب تمرکز داشته‌اند. زیرا انتشار چند برچسب پیچیدگی محاسباتی را افزایش می‌دهد. از این‌رو عمدۀ چالش این مقاله انتشار چند برچسب روی گراف است. الگوریتم انتشار برچسب چندگانه مطرح شده در این مقاله یک‌رونده تکرارشونده است. عملیات انتشار برچسب و شرح‌گذاری در دو مرحله استنتاج و تلفیق انجام می‌شود. در مرحله استنتاج در هر تکرار عملیات انتشار برچسب بر روی هر گراف صورت می‌گیرد. سپس در انتهای تکرار، برچسب‌های به دست آمده برای تصاویر متناظر در دو گراف، در لایه تلفیق با یکدیگر تلفیق خطی شده و مجدد روی تصاویر هر دو گراف انتشار می‌یابد. این فرآیند تکرارشونده تا رسیدن به

<sup>1</sup> Multi-Class

<sup>2</sup> Correlation

یک همگرایی ادامه می‌یابد. پس از اتمام فرآیند شرح‌گذاری، برای افزایش دقت، یک روش پالایش تگ نیز روی نتایج به دست آمده اعمال می‌شود. این روش، تگ‌های ناکامل و یا دارای املای اشتباه را حذف می‌کند.

در [۲۶] یک روش یادگیری نیمه‌نظراتی مبتنی بر دامنه متقابل<sup>۱</sup> به منظور شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ ارائه شده است. این روش برای شرح‌گذاری یک مجموعه تصویر هدف شامل تصاویر برچسبدار و بدون برچسب از مجموعه دیگری بنام دامنه کمکی<sup>۲</sup> استفاده می‌کند. به ازای هر طبقه موجود در مجموعه هدف، یک تابع بازالت گذرانده می‌شود. در این تابع، هدف کمینه کردن فاصله دو توزیع احتمالی تصاویر موجود در دامنه هدف و دامنه کمکی است. در این مقاله برای حل مسئله بهینه‌سازی فرموله شده از روش بیشینه میانگین اختلاف<sup>۳</sup> استفاده شده است.

## ۴-۴- توزیع‌پذیری محاسبات

در برخی از روش‌های ارائه شده برای حل چالش مقیاس‌پذیری از رویکرد محاسبات توزیع‌شده استفاده شده است [۱۰، [۲۸]، [۲۹] و [۲۵]]. در این روش‌ها از مدل‌های یادگیری بکار رفته در مجموعه تصاویر با مقیاس متوسط و کم استفاده می‌شود. ایده اصلی مقاله‌ها برای غلبه بر چالش‌های مطرح در مجموعه تصاویر با مقیاس بزرگ، استفاده از بسترها توزیع‌پذیری محاسبات بر روی کامپیوترها در یک شبکه توزیع‌شده می‌باشد.

در [۲۸] یک قاب‌کاری برای شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ ارائه شده است. این مقاله برای غلبه بر چالش زمان شرح‌گذاری با وجود تعداد زیاد تصاویر و کلمه‌های شرح‌گذاری از رویکرد محاسبات توزیع‌شده استفاده کرده است. ایده این مقاله استخراج ویژگی از تصاویر به صورت موازی و بر روی یک بستر توزیع‌شده و بکارگیری یک الگوریتم SVM سریع در حل مسئله طبقه‌بندی است. از آنجایی که فرآیند استخراج ویژگی از هر تصویر مستقل از یکدیگر

<sup>1</sup> Cross-Domain

<sup>2</sup> Auxiliary Domain

<sup>3</sup> Maximum Mean Discrepancy(MMD)

است، این مقاله این فرآیند را به صورت توزیع شده و با بکارگیری شیوه برنامه نویسی نگاشت، کاهش<sup>۱</sup> انجام داده است. این فرآیند بر روی بستر توزیع شده هادوپ پیاده سازی شده است. برای یادگیری مدل به منظور شرح گذاری از طبقه بند SVM استفاده شده است. مدل SVM استفاده شده از روش شبیه نزولی مبتنی بر میانگین تصادفی<sup>۲</sup> استفاده می کند. این روش به دلیل اینکه تکرار کمتری تا همگرایی انجام می دهد، برای داده های با مقیاس بزرگ مناسب است. درنهایت نیز برای آموزش تعداد زیاد طبقه بند از ایده استفاده از چندین برنامه و به اشتراک گذاری حافظه در بین آنها استفاده کرده است.

در [۱۰] و [۲۵] نیز برای حل چالش زمان شرح گذاری در مجموعه تصاویر در مقیاس بزرگ از راهکاری مشابه یعنی توزیع پذیری محاسبات استفاده شده است. ایده اصلی این روش ها ساخت گراف ها و انتشار برچسب به شیوه برنامه نویسی نگاشت، کاهش و بر روی بستر توزیع شده هادوپ است.

## ۵-۲- جمع بندی

مهمنترین چالش های پیش روی سیستم های شرح گذاری خودکار مجموعه تصاویر در مقیاس بزرگ مقیاس پذیری سیستم، زمان شرح گذاری و دقت شرح گذاری هستند. تابحال پژوهش های گسترده ای برای حل هر یک از این چالش ها صورت گرفته است. در این فصل پژوهش های صورت گرفته در این حوزه مورد بررسی قرار گرفته و دسته بندی ای از آنها ارائه شد. دسته بندی ارائه شده بر اساس رویکرد هر پژوهش در حل هر یک از چالش های مطرح شده می باشد. این دسته بندی شامل روش های پیش پردازشی، مدل های یادگیری و توزیع پذیری محاسبات هستند. مروری بر پژوهش های بررسی شده در هر یک از دسته بندی ها در جدول ۱-۲ آورده شده است.

<sup>1</sup>Map Reduce

<sup>2</sup>Averaging stochastic gradient descent

جدول ۱-۲) خلاصه پژوهش‌های بررسی شده در حوزه شرح‌گذاری خودکار مجموعه تصاویر در مقیاس

بزرگ.

شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ	پیش‌پردازش داده‌ها	بکارگیری الگوریتم ژنتیک مبتنی بر کروموزوم‌های کد شده‌ی دودویی در انتخاب ویژگی برای شرح‌گذاری تصاویر در مقیاس بزرگ.	لی و همکاران [۱۰]
بکارگیری روش انتخاب ویژگی پراکنده مبتنی بر نرم ماتریسی $L_{2,1}$ همچنین بکارگیری زیر فضای به اشتراک گذاشته شده مابین برچسب‌ها در قاب کاری ارائه شده برای شرح‌گذاری خودکار تصاویر در مقیاس بزرگ.	ما و همکاران [۱۸]	بکارگیری روش انتخاب ویژگی پراکنده مبتنی بر نرم ماتریسی $L_{2,1/2}$ در قاب کاری ارائه شده برای شرح‌گذاری تصاویر در مقیاس بزرگ.	شی و همکاران [۱۷]
هش کردن ویژگی‌ها و کاهش ابعاد بردار ویژگی.	ژانگ و همکاران [۱۹]	هش کردن ویژگی‌ها و کاهش ابعاد بردار ویژگی.	کان و همکاران [۲۰]
ارائه توابع هش دودویی در تبدیل بردار ویژگی فضای اصلی به فضای کاهش یافته.	دنگ و همکاران [۲۱]	هش محلی ویژگی‌ها در مقیاس پذیر کردن گراف‌ها.	چن و همکاران [۲۲]
انتخاب پیش‌نمونه‌ها و ساخت گراف پیش‌نمونه‌ها در یادگیری نیمه‌نظرارتی مبتنی بر گراف.	امیری و همکاران [۲۳]	ارائه یک روش KNN وزن‌دار و استفاده از سطوح WordNet برای شرح‌گذاری تصاویر در مقیاس بزرگ.	که و همکاران [۲]
ارائه یک مفهوم جدید ناحیه همسایگی محلی و بکارگیری طبقه‌بند SVM برای شرح‌گذاری تصاویر در مقیاس بزرگ.	لی و همکاران [۱۰]	بکارگیری یادگیری نیمه‌نظرارتی مبتنی بر گراف و ساختار اشتراکی موجود مابین برچسب‌ها برای شرح‌گذاری تصاویر در مقیاس بزرگ.	يانگ و همکاران [۴]
ارائه یک راهکار جدید انتشار چند برچسب در یادگیری نیمه‌نظرارتی مبتنی بر گراف برای شرح‌گذاری تصاویر در مقیاس بزرگ.	لی و همکاران [۲۵]	استفاده از یک مجموعه تصویر کمکی و ارائه یک روش یادگیری نیمه‌نظرارتی مبتنی بر دامنه متقابل در شرح‌گذاری تصاویر در مقیاس بزرگ.	يوآن و همکاران [۲۶]

استخراج ویژگی موازی در یک بستر توزیع شده و بکارگیری یک الگوریتم SVM سریع برای شرح‌گذاری تصاویر در مقیاس بزرگ.	لین و همکاران [۲۸]	توزیع پذیری محاسبات
ساخت بردار ویژگی‌های هر تصویر بصورت موازی و در یک بستر توزیع شده.	لی و همکاران [۱۰]	
توزیع پذیری ساخت گراف و انتشار برچسب در یک بستر توزیع شده برای شرح‌گذاری تصاویر در مقیاس بزرگ.	لی و همکاران [۲۵]	

# فصل ۳: مروی بر روش‌های انتخاب نمونه‌های اولیه

### ۱-۳- پیشگفتار

همانطور که در فصل اول بیان شد، مقیاس‌پذیری و زمان شرح‌گذاری از چالش‌های مهم پیش‌روی سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ هستند. از طرفی، در فصل دوم پژوهش‌های صورت گرفته در حل این چالش‌ها با رویکرد پیش‌پردازش داده‌ها مورد بحث و بررسی قرار گرفت. یکی از رویکردهای پیش‌پردازش داده‌ها که تاکنون در حوزه شرح‌گذاری خودکار تصاویر در مقیاس بزرگ مورد استفاده قرار نگرفته است، انتخاب نمونه‌ها است. در انتخاب نمونه‌ها هدف اصلی یافتن زیرمجموعه‌ای کمینه و تا حد امکان بهینه و با حجمی کمتر از داده‌های اصلی است. این زیرمجموعه شامل نمونه‌های موثر در فرآیند طبقه‌بندی بوده و از آن به عنوان مجموعه آموزش جدید روش‌های یادگیری استفاده می‌شود. با بکارگیری زیرمجموعه آموزش کاهش یافته در روش‌های یادگیری، زمان یادگیری و همچنین زمان طبقه‌بندی کاهش می‌یابد. با کاهش زمان یادگیری مقیاس‌پذیری مدل یادگیری تضمین شده و با مقیاس‌پذیر شدن مدل یادگیری، زمان طبقه‌بندی کاهش می‌یابد. از طرفی با انتخاب نمونه‌های مفید و موثر در طبقه‌بندی دقیق طبقه‌بندی بهبود می‌یابد. بنابراین با بکارگیری این ایده سه چالش اصلی مورد بحث در سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ حل می‌شوند.

همانطور که بیان شد، ایده اصلی این پایان‌نامه انتخاب نمونه‌های موثر در فرآیند یادگیری است. از طرفی روش‌های بکارگرفته شده در جهت حل چالش‌های مطرح شده جز روش‌های PS هستند. از این‌رو در این فصل دسته بندی‌ای از روش‌های PS ارائه شده و این روش‌ها به صورت اجمالی معرفی خواهند شد. همچنین دو روش PS پایه‌ای مورد استفاده در این پایان‌نامه مورد بررسی قرار خواهند گرفت.

### ۲-۳- انتخاب نمونه

IS یکی از روش‌های مقیاس‌پذیر کردن سیستم‌های یادگیری است. فرض اصلی مقیاس‌پذیری در این فرآیند، یافتن نمونه‌های موثر و حذف نمونه‌های بی‌تأثیر در فرآیند طبقه‌بندی و در نتیجه کاهش حجم مجموعه آموزش روش‌های یادگیری است. روش‌های IS به دو

دسته انتخاب مجموعه آموزش<sup>۱</sup> (TSS) و انتخاب نمونه‌های اولیه (PS) تقسیم می‌شوند. PS فرآیند یافتن نمونه‌های نماینده هر طبقه و حذف نمونه‌های بی تاثیر در فرآیند طبقه‌بندی است. در انتخاب نمونه‌ها به وسیله روش‌های PS از قاعده نزدیکترین همسایه استفاده می‌شود.

از طرف دیگر TSS فرآیند یافتن مدل یادگیری مناسب در یک مجموعه آموزش است. در این روش‌ها براساس مدل یادگیری موجود، نمونه‌های مناسب یافت می‌شوند. هر دو دسته روش‌های PS و TSS با استفاده از رویکردی سعی در یافتن زیرمجموعه آموزش مناسب دارند که دارای حجمی کمتر از زیرمجموعه اصلی است. ولی روش‌های TSS فقط در بهبود دقت مدل‌های یادگیری از قبیل بوستینگ<sup>۲</sup>، شبکه‌های عصبی مصنوعی<sup>۳</sup> و درختان تصمیم<sup>۴</sup> بر روی یک مجموعه مرکز داشته و توجهی به تعداد نمونه‌های انتخاب شده ندارند.

روش‌های PS نسبت به TSS دو مزیت عمده دارند. اول اینکه روی تعداد نمونه‌های انتخاب شده توجه داشته و از این‌رو در کاربردهای مقیاس‌پذیری با ایده پیش‌پردازش داده و انتخاب نمونه‌های موثر، مناسب هستند. و دوم اینکه مراحل یادگیری کمتری دارند. به عبارت دقیق‌تر فرآیند یادگیری در روش‌های TSS، زیاد، زمان بر و در مواردی تکرار شونده است. از این‌رو استفاده از روش‌های TSS در کاربردهای مقیاس بالا مناسب نیست [۳۰]. در این پایان‌نامه نیز از روش‌های PS استفاده شده است. از این‌رو در قسمت بعد ویژگی‌های اصلی و طبقه‌بندی کلی از این روش‌ها ارائه می‌شود.

### ۳-۳- انتخاب نمونه‌های اولیه

انتخاب نمونه‌های اولیه (PS) فرآیند یافتن زیرمجموعه‌ای بهبینه و کمینه از داده‌های اصلی با استفاده از قاعده نزدیکترین همسایه‌ها است. زیر مجموعه کاهش یافته به عنوان مجموعه آموزش جدید در روش‌های یادگیری مورد استفاده قرار می‌گیرد. دسته‌بندی روش‌های PS در [۳۰] و [۳۱]<sup>۵</sup> ارائه شده است.

<sup>1</sup> Training Set Selection (TSS)

<sup>2</sup> Boosting

<sup>3</sup> Artificial Neural Networks

<sup>4</sup> Decision Trees

روش‌های PS بر اساس رویکردهای مختلفی دسته‌بندی می‌شوند. این رویکردها شامل نحوه بدست آوردن زیرمجموعه نمونه‌های اولیه، نحوه جستجوی نمونه‌های اولیه و ارزیابی جستجو می‌باشد [۳۱]. روش‌های PS بر اساس رویکرد بدست آوردن زیرمجموعه نمونه‌های اولیه به سه دسته کلی تنکسازی<sup>۱</sup>، اصلاح‌سازی<sup>۲</sup> و ترکیبی<sup>۳</sup> تقسیم‌بندی می‌شوند. هدف روش‌های تنکسازی، حفظ نمونه‌های نزدیک به مرزهای طبقه‌بندی و حذف نمونه‌های دورتر از این مرزها است. از معروف‌ترین روش‌های تنکسازی می‌توان به CNN [۳۲]، MCNN [۳۳]، RNN [۳۴] و FCNN [۳۵] اشاره کرد.

تمرکز اصلی روش‌های اصلاح‌سازی، بهبود دقت طبقه‌بندی است. این روش‌ها با حذف نمونه‌های دارای اختلال سعی دارند با کاهش حجم مجموعه آموزش دقت طبقه‌بندی و رگرسیون را افزایش دهند. از معروف‌ترین روش‌های اصلاح‌سازی می‌توان به ENN [۳۶]، AllKNN [۳۷]، ENN [۳۶] و MultiEdit [۳۸] و RNGE [۳۹] اشاره کرد.

در نهایت روش‌های ترکیبی با بهره‌گیری از ایده دو دسته قبل سعی در یافتن زیرمجموعه کمینه نمونه‌های اولیه به همراه افزایش دقت طبقه‌بندی دارند. از معروف‌ترین روش‌های ترکیبی می‌توان به ICF [۴۰] و Drop3 [۴۱] اشاره کرد.

دومین رویکرد دسته‌بندی، نحوه جستجوی نمونه‌های اولیه است. روش‌های PS بر اساس این رویکرد به پنج گروه افزایشی<sup>۴</sup>، کاهشی<sup>۵</sup>، دسته‌ای<sup>۶</sup>، تلفیقی<sup>۷</sup> و ثابت<sup>۸</sup> تقسیم‌بندی می‌شوند. در رویکرد افزایشی، فرآیند انتخاب نمونه‌ها با زیر مجموعه خالی شروع می‌شود. در هر مرحله نمونه‌ای که دارای شرایط خاص هر روش باشد به زیر مجموعه اضافه می‌شود. از آنجایی که انتخاب هر نمونه بصورت تصادفی است، روش‌های دارای این رویکرد ترتیب محور<sup>۹</sup> هستند [۳۱]. روش‌های

<sup>1</sup> Condensation

<sup>2</sup> Edition

<sup>3</sup> Hybrid

<sup>4</sup> Incremental

<sup>5</sup> Decremental

<sup>6</sup> Batch

<sup>7</sup> Mixed

<sup>8</sup> Fixed

<sup>9</sup> Order dependent

ترتیب محور دارای مراحل یادگیری تکراری و سرعت پایین می‌باشد. از جمله معروفترین روش‌های دارای این رویکرد می‌توان به CNN [۳۲]، MCNN [۳۵]، IB3 [۱۱] و [۴۲] اشاره کرد.

در رویکرد کاهشی، فرآیند انتخاب نمونه‌ها با زیر مجموعه‌ای شامل تمام نمونه‌های آموزش آغاز می‌شود. در هر مرحله، نمونه‌ای که دارای شرایط خاص تعریف شده باشد، از زیرمجموعه حذف می‌شود. از جمله روش‌های دارای این رویکرد می‌توان به RNN [۳۴]، MultiEdit [۳۸] و [۳۹] RNGE اشاره کرد.

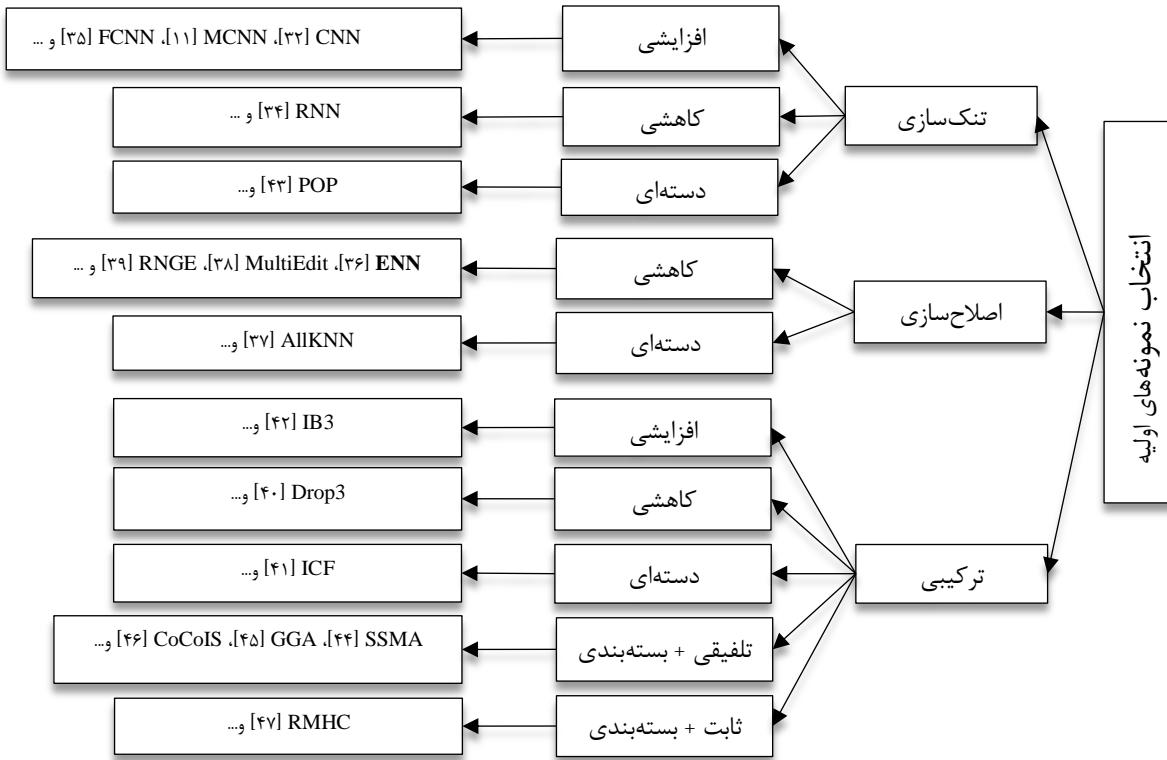
در روش‌های دسته‌ای حذف نمونه‌ها به صورت دسته‌ای بوده و فرآیند حذف چندین نمونه در یک زمان معین صورت می‌گیرد. POP [۴۳] و AllKNN [۳۷] از معروف‌ترین روش‌های دارای این رویکرد می‌باشد.

تلفیقی و ثابت دو رویکرد دیگر نحوه جستجوی نمونه‌های اولیه است. روش‌های دارای رویکرد تلفیقی با یک زیرمجموعه تصادفی شامل چندین نمونه شروع به کار می‌کنند و در هر مرحله نمونه‌های دارای شرایط خاص می‌توانند به زیر مجموعه کاهش یافته اضافه شوند. روش‌های SSMA [۴۴]، GGA [۴۵] و CoCoIS [۴۶] از این رویکرد استفاده می‌کنند. در نهایت در رویکرد ثابت که زیرمجموعه تلفیقی می‌باشد، تعداد حذف و اضافه‌ها مشابه و در ابتدای مرحله اجرایی روش مشخص می‌شود. معروف‌ترین روش دارای این رویکرد، RMHC [۴۷] است.

در نهایت روش‌های PS از لحاظ ارزیابی جستجوی نمونه‌ها با استفاده از قاعده نزدیکترین همسایه‌ها به دو دسته فیلتر<sup>۱</sup> و بسته‌بندی<sup>۲</sup> تقسیم‌بندی می‌شوند. در فیلتر قاعده نزدیکترین همسایه‌ها روی زیرمجموعه جزئی از داده‌ها اعمال می‌شود. در صورتی که در بسته‌بندی فرآیند یافتن نمونه‌های اولیه روی کل مجموعه داده‌هاست. شکل ۱-۳ خلاصه‌ای از طبقه‌بندی روش‌های PS مرور شده در این قسمت را نشان می‌دهد.

<sup>1</sup> Filter

<sup>2</sup> Wrapper



شکل ۱-۳: طبقه‌بندی روش‌های PS موروث شده در این پایان‌نامه.

#### ۴-۳ - روش تنکسازی مبتنی بر نزدیکترین همسایه (CNN)

روش تنکسازی نزدیکترین همسایه<sup>۱</sup> (CNN) در سال ۱۹۶۸ توسط هارت<sup>۲</sup> ارائه شد [۳۲].

فرض کنید  $[l_i \in R^d | 1 \leq i \leq n]$  مجموعه آموزش می‌باشد که در آن  $n$  تعداد  $(n)$ , بردار ویژگی دیداری سطح پایین محاسبه شده برای نمونه  $i$ ام مجموعه آموزش و  $n$  تعداد نمونه‌های مجموعه آموزش می‌باشد. همچنین  $Y = [Y_1, Y_2, Y_3, \dots, Y_n] \in \{0,1\}^{n*c}$ ، ماتریس طبقه نمونه‌های آموزش می‌باشد.  $c$  تعداد طبقه‌ها و  $(Y_i \in R^c | 1 \leq i \leq n)$  بردار طبقه نمونه  $i$ ام است.  $X_{test} = [u_1, u_2, u_3, \dots, u_m]$  مجموعه آزمایش شامل  $m$  نمونه بدون طبقه می‌باشد. روش CNN با استفاده از رویکردی حریصانه سعی در یافتن زیرمجموعه سازگار  $S \subseteq X_{train}$  دارد. الگوریتم CNN با مجموعه تهی  $S$  شروع به کار می‌کند. در هر گذر، یک نمونه آموزش  $i$  از  $X_{train}$  به صورت تصادفی و با شرط نامساوی بودن طبقه  $i$  با طبقه  $Y_i$  با طبقه نزدیکترین همسایه آن در مجموعه  $S$  به این مجموعه اضافه می‌شود. این روند به صورت تکراری و تا رسیدن به همگرایی و عدم تغییر  $S$

<sup>1</sup> Condensed Nearest Neighbor (CNN)

<sup>2</sup> Hart

در یک تکرار ادامه می‌یابد [۳۲]. در این روش از نزدیکترین همسایه هر نمونه استفاده می‌شود در صورتی که می‌توان از K نزدیکترین همسایه هر نمونه استفاده کرد [۳۱].

از مهم‌ترین ایرادات این روش می‌توان به ترتیب محور<sup>۱</sup> بودن و حساس بودن به نمونه‌های پرت و دارای اختلال<sup>۲</sup> اشاره کرد. همچنین این روش تضمینی در یافتن زیرمجموعه سازگار بهینه نمی‌دهد. در سال‌های اخیر پژوهش‌های متعددی از جمله [۳۵] و [۴۸]<sup>۳</sup> سعی در رفع ایرادات این روش و ارائه نسخه‌های بهبودیافته از این روش کرده‌اند. در [۴۸] مشکل ترتیب محور بودن روش CNN حل شده است و نسخه‌ای تغییریافته از این روش (MCNN) ارائه شده است. در [۳۵] به سرعت همگرایی روش CNN در مجموعه‌های بزرگ پرداخته و نسخه‌ای سریع از آن ارائه شده است (FCNN). در رویکرد اول ارائه شده این پایان‌نامه در انتخاب نمونه‌های اولیه نیز به منظور دستیابی به زیرمجموعه سازگار کمینه، با در نظر گرفتن میزان پرت بودن طبقه‌ای یک نمونه در فرآیند تصمیم‌گیری در روش CNN، نسخه بهبودیافته‌ای از آن ارائه شده است.

### ۵-۳- روش اصلاح‌سازی مبتنی بر نزدیکترین همسایه (ENN)

روش اصلاح‌سازی نزدیکترین همسایه<sup>۴</sup> (ENN) [۳۶] در سال ۱۹۷۲ توسط ویلسون<sup>۵</sup> ارائه شد. این روش از دسته روش‌های اصلاح‌سازی است و از رویکرد جستجوی کاهشی استفاده می‌کند. همانطور که در بخش ۳-۳ بیان شد، هدف عمدۀ روش‌های اصلاح‌سازی حذف نمونه‌های دارای اختلال در مجموعه، افزایش دقت و کاهش حجم همزمان است.

روش ENN فرآیند یافتن زیرمجموعه  $S$  را با  $S = X_{train}$  با آغاز می‌کند. به ازای هر نمونه  $j$  از  $K$  در صورتی که بردار برچسب  $Y_j$  با بردار برچسب  $X_{train}$  حداکثری  $K$  نمونه همسایه‌اش برابر نباشد، حذف می‌شود. این روند یک بار و به ازای تمام نمونه‌های آموزش انجام می‌شود. در انتهای این فرآیند،  $S$  شامل نمونه‌های اولیه است و می‌توان از آن به عنوان مجموعه آموزش کاهش یافته جدید در فرآیند یادگیری استفاده کرد.

<sup>1</sup> Order dependent

<sup>2</sup> Edited Nearest Neighbor (ENN)

<sup>3</sup> Wilson

۳-۶- جمع‌بندی

IS فرآیندی پیش‌پردازشی در جهت انتخاب نمونه‌های موثر در یادگیری است. این فرآیند به دو دسته روش‌های TSS و PS تقسیم بندی می‌شود. مزیت روش‌های PS نسبت به TSS سرعت بالاتر و مراحل یادگیری کمتر می‌باشد. در این فصل روش‌های PS ارائه شده مورد بررسی قرار گرفت و طبقه‌بندی از ویژگی‌های آنها ارائه شد. همچنین روند انتخاب نمونه‌های اولیه در دو روش PS مورد استفاده در این پایان‌نامه مورد بررسی قرار گرفت. در ادامه این پایان‌نامه و در فصل بعد به رویکردهای پیشنهادی در انتخاب نمونه‌های اولیه خواهیم پرداخت.

# فصل ۴:

## رویکردهای پیشنهادی

## ۱-۴ - پیشگفتار

همانطور که در فصل اول بیان شد، شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ فرآیندی چند مرحله‌ای است. این فرآیند شامل پیش‌پردازش، شرح‌گذاری و پس‌پردازش می‌باشد. تمرکز اصلی این پایان‌نامه بر روی پیش‌پردازش مجموعه آموزش و انتخاب نمونه‌های اولیه می‌باشد. از این‌رو در ابتدا و در بخش ۲-۴ دو رویکرد پیشنهادی این پایان‌نامه در انتخاب نمونه‌های اولیه تشریح می‌شود. به منظور سنجش نحوه انتخاب نمونه‌های اولیه و بهبود دقت شرح‌گذاری، دو الگوریتم یادگیری به کار گرفته شده است. نحوه شرح‌گذاری تصاویر بدون برچسب توسط این دو الگوریتم و همچنین نحوه تصمیم‌گیری چندبرچسبه در فرآیند شرح‌گذاری نیز در بخش ۳-۴ مورد بحث و بررسی قرار گرفته است.

## ۲-۴ - انتخاب نمونه‌های اولیه

در این بخش دو رویکرد پیشنهادی این پایان‌نامه در انتخاب نمونه‌های اولیه تشریح می‌شود. رویکرد اول بهبود عملکرد روش CNN با ایده حذف نمونه‌های پرت طبقه‌ای است. رویکرد دوم نیز بهبود عملکرد و کاربردی سازی روش ENN در کاربردهای چندبرچسبه و ارائه الگوریتمی نو در انتخاب نمونه‌های اولیه است.

## ۱-۲-۴ رویکرد پیشنهادی اول در انتخاب نمونه‌های اولیه

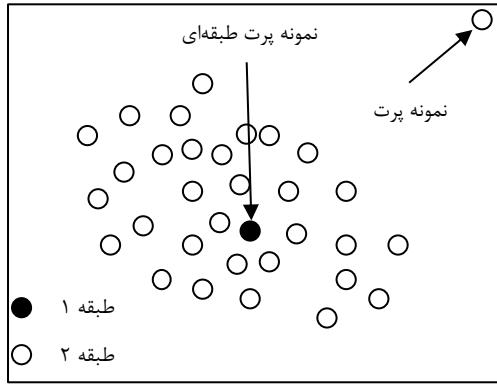
رویکرد پیشنهادی اول این پژوهش، روشی ترکیبی به منظور یافتن زیرمجموعه سازگار کمینه است. ایده اصلی این روش در نظر گرفتن و حذف نمونه‌های پرت طبقه‌ای در روش پایه‌ای CNN است، که تاکنون در دیگر پژوهش‌های مبتنی بر تنکسازی به آن پرداخته نشده است. بکارگیری این ایده منجر به بهبود عملکرد روش CNN در یافتن زیرمجموعه سازگار کمینه می‌شود. از این‌رو در ابتدا ایده حذف نمونه‌های پرت طبقه‌ای بکار رفته در روش CNN و سپس رویکرد پیشنهادی مطرح خواهد شد.

#### ۱-۱-۲-۴ نمونه‌های پرت طبقه‌ای

نمونه پرت، نمونه‌ای است که در میان نمونه‌ها دارای رفتاری متفاوت است. در بیشتر کاربردهای تشخیص الگو و داده‌کاوی نمونه‌های پرت اختلال محسوب شده و از این‌رو شناسایی و حذف آن‌ها از اهمیت بالایی برخوردار است. روش‌های متعددی نیز از جمله روش‌های آماری، خوشبندی، مبتنی بر فاصله و مبتنی بر مدل یادگیری برای تشخیص این نمونه‌ها ارائه شده است.

تمام این روش‌ها فرآیند تشخیص را بدون توجه به طبقه نمونه‌ها انجام می‌دهند [۴۹].

با در نظر گرفتن طبقه نمونه‌ها، در [۴۹] نوع دیگری از نمونه‌های پرت، بنام نمونه‌های پرت طبقه‌ای مطرح شده است. همان‌طور که قبلاً نیز بیان شد، نمونه پرت طبقه‌ای، نمونه‌ای است که در بین گروهی از نمونه‌های با طبقه مشابه قرار گرفته و دارای طبقه متفاوت با آن‌ها می‌باشد. شکل ۱-۴ تفاوت مابین نمونه‌های پرت و نمونه‌های پرت طبقه‌ای را نشان می‌دهد.



شکل ۱-۴: تفاوت مابین نمونه پرت و پرت طبقه‌ای.

در [۴۹] به منظور تشخیص نمونه‌های پرت طبقه‌ای، راهکاری مبتنی بر فاصله ارائه شده است. در این راهکار با استفاده از رابطه (۱-۴) میزان پرت طبقه‌ای<sup>۱۴</sup> بودن ( $COF$ ) نمونه  $T$  بدست می‌آید. در این رابطه  $K$ ، تعداد نزدیک‌ترین همسایه‌های نمونه  $T$  و  $PCL(T, K)$  احتمال طبقه نمونه  $T$  در بین طبقه  $K$  نزدیک‌ترین همسایه‌هایش می‌باشد.

$$COF(T) = K * PCL(T, K) + \alpha * \frac{1}{Dev(T)} + \beta * KDis(T) \quad (1-4)$$

برای مثال اگر در شکل ۱-۴ مقدار  $K$  را برابر ۷ فرض کنیم، مقدار  $PCL$  نمونه طبقه ۱ برابر ۱/۷ است. با فرض داشتن مجموعه  $DCL = [l_1, l_2, l_3, \dots, l_h]$  شامل  $h$  نمونه با طبقه مشابه با نمونه  $T$  که  $DCL \subseteq X_{train}$  در این رابطه  $Dev(T)$  میزان انحراف نمونه  $T$  از این مجموعه است. این مقدار به صورت مجموع فواصل نمونه  $T$  از نمونه‌های  $DCL$  تعریف می‌شود. همچنین در این رابطه  $KDis(T)$  بیانگر میزان نزدیکی نمونه  $T$  به  $K$  نزدیک‌ترین همسایه‌هایش می‌باشد. ضرایب  $\alpha$  و  $\beta$  به منظور کنترل اهمیت و تأثیر مقادیر  $Dev$  و  $KDis$  در این رابطه گنجانده شده است و مقادیر آن‌ها به حوزه کاربرد وابسته بوده و به صورت تجربی تعیین می‌شود. همچنین با ضرب  $PCL$  در  $K$ ، مقدار آن از بازه  $[I, K]$  به  $[I/K, I]$  تغییر می‌یابد. درنهایت نمونه‌ای با مقدار  $COF$  کمتر، دارای درجه پرتوی بیشتر است. روابط (۲-۴) و (۳-۴) به ترتیب فرمول محاسبه  $Dev(T)$  و  $KDis(T)$  را نشان می‌دهند. در این روابط  $Dis$  بیانگر فاصله مابین دو نمونه است [۴۹].

$$Dev(T) = \sum_{i=1}^h Dis(T, l_i) \quad l_i \in DCL \quad (2-4)$$

$$KDis(T) = \sum_{i=1}^K Dis(T, l_i) \quad (3-4)$$

## ۲-۱-۲-۴ رویکرد پیشنهادی COCNN

رویکرد تنکسازی نزدیک‌ترین همسایه مبتنی بر حذف نمونه‌های پرت طبقه‌ای<sup>۱</sup> (COCNN) ارائه شده در این قسمت، از ایده تشخیص نمونه‌های پرت طبقه‌ای ارائه شده در [۴۹] و ترکیب آن با روش CNN استفاده می‌کند. همان‌طور که قبلاً بیان شد، در صورت وجود نمونه‌هایی با طبقه متفاوت در فاصله نزدیک به نمونه‌های با طبقه دیگر، روش CNN در یافتن زیرمجموعه سازگار کمینه دچار مشکل می‌شود. از این‌رو نوآوری اصلی COCNN حل این مشکل و بهبود عملکرد روش CNN، با بکارگیری روش تشخیص نمونه‌های پرت طبقه‌ای در فرآیند تصمیم‌گیری تعلق یا عدم تعلق یک نمونه به زیرمجموعه سازگار می‌باشد.

فرآیند تصمیم‌گیری در الگوریتم COCNN بدین صورت است که یک نمونه در صورت دارا بودن دو شرط زیر به زیرمجموعه سازگار اضافه می‌شود: ۱) دارای طبقه‌ای نامساوی با طبقه

<sup>۱</sup> Class Outlier based Condensed Nearest Neighbor (COCNN)

نزدیک‌ترین همسایه خود در  $S$  باشد و ۲) دارای درجه پرتوی کمتری نسبت به نزدیک‌ترین همسایه خود در  $S$  باشد. به عبارت دیگر دارای مقدار  $COF$  بیشتری نسبت به  $COF$  نزدیک‌ترین همسایه خود در  $S$  باشد. الگوریتم ۱-۴ شبیه کد روش COCNN ارائه شده را نشان می‌دهد.

---

#### COCNN algorithm

---

**Inputs:**

$X_{train} = [l_1, l_2, l_3, \dots, l_n]$ , Training instances  
 $Y = [Y_1, Y_2, Y_3, \dots, Y_n]$ , Training labels

**Output:**

$S \subseteq X_{train}$

**Process:**

```

 $S \leftarrow \emptyset$ 
foreach ( $l_i \in X_{train}$ )
    Compute  $COF(l_i)$  by using equation (4-1)
end
Repeat until convergence
    foreach ( $l_i \in X_{train}$ ) (in random order)
        find nearest neighbor of  $l_i$  ( $l_j$ ) in  $S$ 
        if ( $y_i \neq y_j$  AND  $COF(l_i) > COF(l_j)$ )
            Add  $l_i$  into  $S$ 
        end
    end
end
Return  $S$ 

```

---

الگوریتم ۱-۴: شبیه کد روش COCNN ارائه شده.

در این الگوریتم ابتدا با استفاده از رابطه (۱-۴) مقدار  $COF$  همه نمونه‌های آموزش بدست می‌آید. سپس بهمانند الگوریتم CNN، در هر گذر، نمونه آموزش  $l_i$  از  $X_{train}$  به صورت تصادفی و با شرط نامساوی بودن طبقه  $y_i$  با طبقه نزدیک‌ترین همسایه‌اش در مجموعه  $S$  و بزرگ‌تر بودن مقدار  $COF(l_i)$  از  $COF(l_j)$  به مجموعه  $S$  اضافه می‌شود. این روند به صورت تکراری و تا رسیدن به همگرایی و عدم تغییر  $S$  در یک تکرار ادامه می‌یابد.

در مجموعه‌هایی که نمونه‌ها با طبقه‌های متفاوت در بین نمونه‌ها با طبقه‌های مشابه قرار گرفته‌اند (نمونه‌های پرت طبقه‌ای) الگوریتم CNN دچار مشکل می‌شود. در این حالت با توجه به شرط تصمیم‌گیری در الگوریتم CNN، این روش به ناچار این نمونه‌ها را به زیرمجموعه سازگار اضافه می‌کند و باعث افزایش حجم این زیرمجموعه می‌شود. دلیل برتری روش COCNN نسبت

به روش CNN در شناسایی این نمونه‌ها و در نظر گرفتن آن‌ها در روند تصمیم‌گیری الگوریتم است. انجام این کار منجر به کاهش حجم زیرمجموعه سازگار با حداقل کاهش دقت می‌شود.

#### ۲-۲-۴ رویکرد پیشنهادی دوم در انتخاب نمونه‌های اولیه

همانطور که بیان شد، مهمترین ایراد روش‌های تنکسازی عدم توجه به دقت، ترتیب محور و زمان بر بودن آنها می‌باشد. هدف اصلی این پایان‌نامه بکارگیری روش‌های PS در کاربردهای مقیاس بزرگ و حل همزمان چالش دقت و مقیاس‌پذیری می‌باشد. روش‌های اصلاح‌سازی در عین سادگی و پیچیدگی محاسباتی کم، توجه ویژه‌ای به دقت طبقه‌بندی دارند. از این‌رو در ادامه این پژوهش روشENN که یکی از روش‌های اصلاح‌سازی است، بکار گرفته شده است. در [۱۲] از روشENN در تعادل سازی مجموعه‌های چندبرچسبه نامتعادل استفاده شده است. از آنجایی که دیگر هدف این پایان‌نامه ارائه و بکارگیری روشPS در طبقه‌بندی چندبرچسبه می‌باشد، از ایده ارائه شده در این مقاله استفاده شده است. بنابراین در ابتدا و در بخش بعد ایده بکار رفته در [۱۲] برای طبقه‌بندی چندبرچسبه ارائه می‌شود. سپس به بیان نوآوری‌های بکار رفته درENN در راستای ارائه روشPS در این پژوهش خواهیم پرداخت. لازم به ذکر است، ایده ارائه شده در [۱۲] اساس کار رویکرد انتخاب نمونه‌های اولیه ارائه شده در این پژوهش می‌باشد.

#### ۱-۲-۲-۴ENN چندبرچسبه

به منظور بهبود عملکرد الگوریتم‌های طبقه‌بندی چندبرچسبه هنگام اعمال بر روی مجموعه‌های چندبرچسبه نامتعادل، روشMLENN ارائه شده است [۱۲]. در این روش برای تعادل‌سازی مجموعه‌ها ازENN استفاده می‌شود. همچنین از آنجایی که به بحث اعمال روش روی داده‌های چندبرچسبه پرداخته شده است، این روش اصلاح‌سازی نزدیکترین همسایه چندبرچسبه (MLENN) نام دارد.

فرض این روش در تعادل سازی مجموعه داده‌های نامتعادل، شناسایی و حفظ نمونه‌هایی است که منجر به وجود آمدن عدم تعادل مجموعه شده است. بدین منظور در ابتدا تمام این

نمونه‌ها شناسایی شده و سپس با اعمال ENN روی مابقی نمونه‌ها، مجموعه‌ها متعادل می‌شوند. نمونه‌هایی که منجر به عدم تعادل مجموعه می‌شوند با دو معیار  $LblIR$  و  $DSIR$  شناسایی می‌شوند. معیار  $LblIR$  نشان دهنده میزان عدم تعادل هر برچسب است. رابطه (۴-۴) نحوه محاسبه این معیار را نشان می‌دهد.

$$LbIIR(y) = \frac{\arg\max(\sum_{i=1}^{\text{size}(D)} k(y', Y_i))}{\sum_{i=1}^{\text{size}(D)} k(y, Y_i)}, \quad k(y', Y_i) = \begin{cases} 1 & \text{if } y \in Y_i \\ 0 & \text{otherwise} \end{cases} \quad (4-4)$$

در این رابطه  $y$  یک برچسب خاص از مجموعه برچسب‌های  $Y$  و  $Y_i$  بردار برچسب نمونه نام است. مقدار  $LbIIR$  بالاتر نشان دهنده میزان عدم تعادل بالاتر است. از طرفی  $DSIR$  نشان دهنده میزان عدم تعادل یک مجموعه داده است و با میانگین گیری از تمام مقادیر  $LbIIR$  بدست می‌آید. رابطه (۵-۴) نحوه محاسبه این معیار را نشان می‌دهد.

$$DSIR = \frac{\sum_{y=y_1}^{y_c} (LbIIR(y))}{\text{Size}(Y)} \quad (5-4)$$

ایده اصلی [۱۲] در تعادل سازی مجموعه‌ها این است که نمونه‌هایی که دارای برچسبی با مقدار  $LbIIR$  کمتر از مقدار  $DSIR$  می‌باشند، منجر به عدم تعادل مجموعه شده و باید حفظ شوند. سپس به منظور تعادل‌سازی، الگوریتم ENN بر روی مابقی نمونه‌ها اعمال شده و نمونه‌هایی که شرط ENN را نداشته باشند، حذف می‌شوند. بدین ترتیب یک مجموعه نامتعادل به یک مجموعه تعادل تبدیل شده و دقت اعمال الگوریتم‌های چندربرچسبه روی مجموعه تعادل افزایش می‌یابد. یکی از ویژگی‌های بارز مجموعه‌های چندربرچسبه مقیاس بزرگ، عدم تعادل زیاد در این مجموعه‌هاست. به عبارت دیگر با افزایش تعداد نمونه‌ها و برچسب‌ها، تعداد نمونه‌های متعلق به یک برچسب کم شده و در نتیجه مجموعه‌ها نامتعادل می‌شوند. از این‌رو به منظور جلوگیری از هرگونه کاهش دقت، در پژوهش جاری از ایده شناسایی و نگهداری نمونه‌های عدم تعادل استفاده شده است.

## ۴-۲-۲-۴ عدم شباهت چندبرچسبی

در مسائل طبقه‌بندی چندبرچسبی به ازای هر نمونه، برداری دودویی که اندیس‌ها نشان دهنده برچسب‌ها است، وجود دارد. در این بردار تعلق و یا عدم تعلق یک برچسب به یک نمونه با صفر و یک نشان داده می‌شود. روش‌های پایه‌ای مورد استفاده در این پژوهش از اصل نزدیکترین همسایه‌ها استفاده می‌کنند. در این روش‌ها در تصمیم‌گیری حضور و یا عدم حضور یک نمونه در زیرمجموعه کاهش یافته برچسب نمونه‌ها با یکدیگر مقایسه می‌شوند.

در نمونه‌های تک برچسبی مقایسه مابین برچسب‌ها دارای دو حالت صفر(مشابه) و ۱(مخالف) است. ولی در نمونه‌های دارای چند برچسب در نظر گرفتن دو حالت شباهت کامل و تفاوت کامل سخت‌گیرانه است و منجر به کاهش کارایی و دقت عملکرد روش می‌شود. پس در نظر گرفتن درجه شباهتی متغیر و مابین ۰ تا ۱ منجر به بهبود عملکرد خواهد شد. برای مثال دو بردار برچسب شکل ۲-۴ را درنظر بگیرید.

	اندیس برچسب								
	۸۱	۸۰		۶	۵	۴	۳	۲	۱
بردار ۱	.	۱	...	.	۱	۱	۰	۰	۱
بردار ۲	.	۱	...	.	۰	۰	۱	۰	۰
تفاوت	.	۰	...	.	۰	۱	۰	۰	۰

شکل ۲-۴: مثالی از عدم شباهت دو بردار برچسب دودویی.

در این شکل دو بردار برچسب با یکدیگر متفاوتند ولی تنها تفاوت آنها در برچسب شماره ۵ است. با بکارگیری شیوه سنتی در مقایسه این دو بردار، عدم تشابه کامل در نظر گرفته می‌شود. در صورتی که این دو بردار برچسب بسیار به یکدیگر شبیه هستند و تنها در یک مورد با یکدیگر تفاوت دارند. آنچه مسلم است در نظر گرفتن عدم تشابه کامل برای آنها کاری اشتباہ است [۱۲].

به منظور محاسبه شباهت و عدم شباهت بردارهای دودویی روابط متعددی ارائه شده است

[۵۰]. فرض کنید  $Y_i$  و  $Y_j$  دو بردار برچسب دودویی دو نمونه  $i$  و  $j$  باشند. همچنین بر اساس تعریف

[۵۰] قرار می‌دهیم:  $A$  تعداد دودویی‌های  $Y_i=1$  و  $B$  تعداد دودویی‌های  $Y_i=0$  و  $C$ ,  $Y_j=0$

تعداد دودویی‌های  $Y_i=0$  و  $Y_j=I$  و  $D$  تعداد دودویی‌های  $Y_i=0$  و  $Y_j=0$ . مقادیر  $A$ ,  $B$ ,  $C$  و  $D$  در شکل ۲-۴ به ترتیب برابر ۳، ۱، ۰ و ۷۷ می‌باشند. برای مثال بر اساس تعاریف ارائه شده در بالا میزان عدمشباخت چندبرچسبی همینگ با استفاده از رابطه  $\frac{B+C}{A+B+C+D}$  محاسبه می‌شود. در مثال شکل ۲-۴ عدمشباخت همینگ برابر ۱۲۳٪ است. همانطور که مشاهده می‌شود عدد بسیار کوچک بدست آمده نشان از شباخت بالای این دو بردار برچسب است.

با بکارگیری عدمشباخت چندبرچسبی همینگ و با وجود تعداد زیاد بردار برچسب که از خواص سیستم‌های شرح‌گذاری مجموعه تصاویر است، این میزان به صفر میل می‌کند و منجر به عدم مقایسه درست می‌شود. برای اجتناب از این شرایط می‌توان در مخرج کسر بجای استفاده از تعداد کل بردار برچسب‌ها، تعداد بردار برچسب‌های ۱ در دو بردار برچسب را قرار داد. با بکارگیری این ایده مقدار عدمشباخت چندبرچسبی بدست آمده قابل مقایسه‌تر خواهد شد. این رابطه به عدمشباخت تنظیم شده<sup>۱</sup> معروف است [۱۲]. در مثال شکل ۲-۴ میزان عدمشباخت چندبرچسبی همینگ تنظیم شده دو بردار برچسب برابر ۱۴۲۹٪ است.

در روش پیشنهادی دوم ارائه شده در این پژوهش، به منظور بررسی دقیق تاثیر بکارگیری روابط عدمشباخت چندبرچسبی مختلف، از روابط محاسبه فاصله دودویی ارائه شده در [۵۰] استفاده شده است. همچنین در برخی از این روابط در صورت امکان از ایده عدمشباخت تنظیم شده استفاده و تاثیر بکارگیری این ایده نیز مورد بررسی قرار گرفته است. از این‌رو دسته دیگری از روابط تغییر یافته ارائه شده است. این روابط به همراه فرمول محاسبه آنها در جدول ۱-۴ لیست شده است. در این جدول به ازای هر رابطه نام اختصار آن نیز آمده است که در ادامه از این نام اختصار استفاده می‌کنیم.

### ۳-۲-۲-۳ پیش‌بینی برچسب از نزدیکترین همسایه‌ها

در تصمیم‌گیری در مورد حذف هر نمونه در روش انتخاب نمونه اولیه ENN، برچسب نمونه کاندیدای حذف و برچسب رای‌دهی شده توسط نزدیکترین همسایه‌های آن با یکدیگر مقایسه

<sup>۱</sup> Adjusted dissimilarity

می‌شوند. بر اساس مقدار عدم شباهت بدست آمده مابین دو برجسب با استفاده از روابط معرفی شده در بخش قبل، تصمیم‌گیری در مورد حذف هر نمونه صورت می‌گیرد. به منظور تصمیم‌گیری درست، بالا بودن کیفیت برجسب‌های رای‌دهی شده از نزدیکترین همسایه‌ها ضروری است. همچنین شمارش عادی و بکارگیری رای‌گیری حداکثری از نزدیکترین همسایه‌ها منجر به عدم صحت رای‌دهی و کیفیت پایین بردار برجسب پیش‌بینی شده می‌شود. پس بکارگیری یک روش پیش‌بینی برجسب مناسب در مرحله تصمیم‌گیری روش ENN منجر به مقایسه درست بردار برجسب نمونه کاندیدای حذف و برجسب رای‌دهی شده توسط نزدیکترین همسایه‌های آن می‌شود.

جدول ۱-۴) روابط عدم شباهت دودویی متعدد.

ردیف	نام	اختصار	رابطه
۱	همینگ	HM	$B + C / A + B + C + D$
۲	همینگ تنظیم‌شده	ADJ-HM	$B + C / Active\ Labels$
۳	تفییرات	VA	$B + C / 4 * (A + B + C + D)$
۴	تفییرات تنظیم‌شده	ADJ-VA	$B + C / 4 * Active\ Labels$
۵	تفاوت اندازه	SD	$(B + C)^2 / (A + B + C + D)^2$
۶	تفاوت اندازه تنظیم شده	ADJ-SD	$(B + C)^2 / (Active\ Labels)^2$
۷	تفاوت شکل	SHD	$(A + B + C + D) * (B + C) - (B - C)^2 / (A + B + C + D)^2$
۸	تفاوت شکل تنظیم شده	ADJ-SHD	$(A + B + C + D) * (B + C) - (B - C)^2 / (Active\ Labels)^2$
۹	تفاوت الگو	PD	$4 * B * C / (A + B + C + D)^2$
۱۰	تفاوت الگوی تنظیم‌شده	ADJ-PD	$4 * B * C / (Active\ Labels)^2$
۱۱	لورانس و ویلیامز	L&W	$(B + C) / (2 * A + B + C)$
۱۲	هلینگر	HELL	$2 * \sqrt{\left(1 - \frac{A}{\sqrt{(A + B)(A + C)}}\right)}$
۱۳	وتری	CH	$\sqrt{2 * \left(1 - \frac{A}{\sqrt{(A + B)(A + C)}}\right)}$

در گام تصمیم‌گیری در مورد حذف هر نمونه در روش پیشنهادی، بجای رای‌گیری حداکثری از نزدیکترین همسایه‌ها، بردار برچسبی از نزدیکترین همسایه‌ها پیش‌بینی می‌شود. در پیش‌بینی بردار برچسب مد نظر از دو روش پیش‌بینی برچسب استفاده شده است. روش اول KELM روشی حریصانه ارائه شده در [۵۱] در کاربرد شرح‌گذاری خودکار تصاویر و روش دوم [۵۲] است. بنابراین در ابتدا و در ادامه این بخش این دو روش را توضیح خواهیم داد.

### روش پیش‌بینی حریصانه

رویکرد پیش‌بینی برچسب حریصانه بکار رفته در این قسمت، روش ارائه شده در [۵۱] است. همانطور که بیان شد، از این الگوریتم به منظور ساخت بردار برچسب نمونه کاندیدای حذف از روی نزدیکترین نمونه‌های آن استفاده می‌شود. نحوه پیش‌بینی  $N$  برچسب برای نمونه کاندیدای حذف توسط این الگوریتم بدین صورت می‌باشد. در ابتدا  $K$  نزدیکترین همسایه‌های نمونه کاندیدای حذف در مجموعه آموزش یافت می‌شود. تعیین  $K$  بهینه از اهمیت بالایی برخوردار است. تنظیم پارامترهای بهینه در فصل بعد بحث و بررسی خواهد شد. در مرحله بعد برچسب‌های نزدیکترین همسایه اول بر اساس تعداد تکرار در مجموعه آموزش (شامل تمام نمونه‌های آموزش)، مرتب می‌شوند. پس از مرتب‌سازی برچسب‌ها بر اساس تعداد تکرار،  $N$  برچسب نزدیکترین همسایه اول دارای بیشترین تعداد تکرار به عنوان بردار پیش‌بینی در نظر گرفته می‌شود. همانطور که بیان شد  $N$  نشان دهنده تعداد برچسب‌های مد نظر برای پیش‌بینی است. مقدار  $N$  با میانگین تعداد برچسب‌های هر نمونه در مجموعه آموزش تعیین می‌شود. در صورتی که تعداد برچسب‌های نزدیکترین همسایه اول با تعداد برچسب تعیین شده به منظور پیش‌بینی برای نمونه کاندیدای حذف، برابر و یا بیشتر از آن باشد، عملیات پیش‌بینی بردار برچسب به پایان می‌رسد. در غیر اینصورت، پیش‌بینی و انتقال برچسب‌های باقیمانده از میان همسایه‌های دوم تا  $K$  صورت می‌گیرد. پیش‌بینی مد نظر از میان سایر همسایه‌ها، بر اساس دو معیار رخداد مشترک<sup>۱</sup> و بسامد<sup>۲</sup> محلی<sup>۳</sup> تعریف شده در زیر انجام می‌شود.

<sup>1</sup> Co-Occurrence

<sup>2</sup> Local Frequency

<sup>3</sup> دسترسی به این مدرک بر پایه آینه‌نامه ثبت و اشاعه پیشنهادهای، پایان‌نامه‌ها، رساله‌های تحصیلات تکمیلی و صیانت از حقوق پدیدآوران در آنها (وزارت علوم، تحقیقات، فناوری به شماره ۱۹۵۹۲۹۰/۹/۶) از پایگاه اطلاعات علمی ایران (گنج) در پژوهشگاه علوم و فناوری اسلامی ایران (ایراندک) فراهم شده و استفاده از آن با رعایت کامل حقوق پدیدآوران و تنها برای هدفهای علمی، آموزشی، پژوهشی و بر پایه قانون حمایت از مؤلفان، مصنفوان، و هنرمندان (۱۳۴۸) و الحالات و اصلاحات بعدی آن و سایر قوانین و مقررات مربوط شدنی است.

- رخداد مشترک: این معیار به صورت رخداد همزمان هر برچسب مد نظر برای پیش‌بینی و برچسب‌های پیش‌بینی شده از مرحله قبل در مجموعه آموزش تعریف می‌شود. مقدار بدست آمده برای این معیار با تقسیم بر مجموع مقادیر، نرمال سازی شده و به عنوان اولین معیار پیش‌بینی در نظر گرفته می‌شود.
  - بسامد محلی: این معیار به صورت تعداد نمونه‌های دارای برچسب مد نظر برای پیش‌بینی در مجموعه نزدیکترین همسایه‌ها تعریف می‌شود. مقدار بدست آمده برای این معیار با تقسیم بر مجموع مقادیر، نرمال سازی شده و به عنوان دومین معیار پیش‌بینی در نظر گرفته می‌شود. پس از بدست آمدن و نرمال شدن مقادیر رخداد مشترک و بسامد محلی و به منظور تشکیل یک معیار واحد برای تصمیم‌گیری، از این دو معیار میانگین‌گیری می‌شود. مقدار میانگین بدست آمده به عنوان معیار کلی برای پیش‌بینی برچسب‌های باقیمانده در نظر گرفته می‌شود.
- در مرحله دوم پیش‌بینی، به تعداد برچسب‌های مورد نیاز (این تعداد برابر  $N$  منهای تعداد برچسب‌های نزدیکترین همسایه)، از میان برچسب‌های همسایه‌های دوم تا  $K$ ام، با شرط مقدار میانگین رخداد مشترک و بسامد محلی بیشتر، پیش‌بینی می‌شود. در انتهای دو مرحله، بردار برچسب نمونه کاندیدای حذف با تعداد  $N$  برچسب بدست می‌آید. استدلال اصلی بکارگیری الگوریتم حریصانه در پیش‌بینی برچسب در این بخش بهبود کیفیت و پیش‌بینی دقیق بردار برچسب نمونه کاندیدای حذف به منظور مقایسه دقیق دو بردار برچسب می‌باشد. الگوریتم پیش‌بینی استفاده شده در این بخش، صرفا بر اساس تکرار برچسب‌ها و کیفیت آنها در مجموعه آموزش فرآیند پیش‌بینی را صورت می‌دهد. از این‌رو پیش‌بینی صورت گرفته از کیفیت بالایی برخوردار است.

### روش پیش‌بینی مبتنی بر الگوریتم KELM

روش ELM [۵۳] و [۵۴] اولین بار به منظور غلبه بر محدودیت‌های بکارگیری روش شبکه عصبی پس‌انتشار<sup>۱</sup> (BP) ارائه شد. دو ویژگی متمایز این روش نسبت به روش یادگیری BP، تعیین

---

<sup>۱</sup> Back Propagation (BP)

تعداد و مقداردهی تصادفی به نرون‌های لایه پنهان می‌باشد. به عبارت دقیق‌تر پارامترهای نرون‌های لایه پنهان در ابتدا و بصورت مستقل از نمونه‌های آموزش مقداردهی شده و فرآیندهای یادگیری تکرار شونده در شبکه‌های عصبی BP به یک فرآیند یادگیری تک مرحله‌ای تبدیل می‌شود. با بکارگیری این رویکرد، محدودیت‌های محاسباتی ELM در مواجه با داده‌های مقیاس بزرگ کاهش قابل توجهی دارد. این روش در سالهای اخیر مورد توجه بسیاری از پژوهشگران حوزه یادگیری قرار گرفته و نسخه‌های متعددی از آن در کاربردهای مختلف از جمله پردازش تصویر و بینایی ماشین [۵۵] تجزیه و تحلیل محتوای ویدیو [۵۶] و پژشکی [۵۷] ارائه و بکار گرفته شده است. نسخه‌های ارائه شده ELM از جمله I-ELM [۵۸]، EM-ELM [۵۹] و OS-ELM [۶۰] سعی در رفع محدودیت‌ها و بکارگیری این روش در کاربردهای مختلف یادگیری داشته‌اند.

با مجموعه آموزش داده شده  $X_{train}$  شامل  $N$  نمونه آموزش، تابع خروجی یک شبکه عصبی با  $L$  نرون لایه پنهان به صورت رابطه (۴-۶) بیان می‌شود.

$$f(X_i) = \sum_{j=1}^L \beta_j h_j(\alpha_j, \beta_j, X_i) = h(X_i) \beta, i = 1 \dots N \quad (\text{Eq-4})$$

در این رابطه  $h(X_i) = [h_1(a_1, b_1, X_i), \dots, h_j(a_L, b_L, X_i)]$  ماتریس وزن‌های لایه خروجی،  $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$  خروجی شبکه بر اساس هر نمونه  $X_i$  و  $(\cdot)$  تابع تبدیل غیر خطی لایه پنهان مانند RBF می‌باشد. همچنین  $a_j$  و  $b_j$  پارامترهای نرون زام لایه پنهان است. فرآیند یادگیری در شبکه شامل کمینه سازی خطای نرم  $\|H\beta - Y\|/2$  است که در آن:

$$H = \begin{bmatrix} h(X_1) \\ \dots \\ h(X_N) \end{bmatrix}, Y = \begin{bmatrix} y_1^T \\ \dots \\ y_N^T \end{bmatrix} \quad (\text{Y-4})$$

در رابطه (۷-۴)،  $H$  و  $Y$  به ترتیب ماتریس خروجی لایه پنهان و ماتریس برچسبها می‌باشد.

در فرآیند یادگیری در شبکه‌های عصبی تک لایه، پارامترهای شبکه به صورت تکرار شونده و با روش گرادیان نزولی<sup>۱</sup> بدست آمده و بروزرسانی می‌شوند. در صورتی که در ELM تعداد نرون‌های لایه پنهان،  $L$  و مقادیر پارامترها به صورت تصادفی مقداردهی می‌شود. به عبارت دقیق‌تر ELM یک رویکرد یادگیری بدون تنظیم<sup>۲</sup> در شبکه‌های عصبی پس انتشار را به اجرا می‌گذارد [۶۱]. از این‌رو فرآیند یادگیری، به حل و بهینه‌سازی یکتابع خطی تبدیل شده و پیچیدگی محاسباتی در آن به شدت کاهش می‌یابد.

رویکرد یادگیری ELM در کمینه‌سازی همزمان خطای یادگیری و نرم ماتریسی وزن‌های خروجی در [۵۲] ارائه شده است. در این رویکرد مسئله بهینه‌سازی به صورت رابطه (۸-۴) تعریف می‌شود.

$$\min \frac{1}{2} \|\beta\|_2^2 + \frac{C}{2} \sum_{i=1}^N \|\xi_i\|^2 \quad s.t. h(X_i)\beta = y_i^T - \xi_i^T \quad i = 1 \dots N \quad (8-4)$$

در رابطه (۸-۴)،  $\xi_i$  و  $y_i$  به ترتیب خطای تخمین و بردار برچسب نمونه آموزش  $X_i$  می‌باشد. همچنین  $C$  ضریب کنترل<sup>۳</sup> است. بر اساس نظریه KKT وزن‌های خروجی  $(\beta)$  با حل بهینه رابطه (۹-۴) بدست می‌آید.

$$L_{ELM} = \frac{1}{2} \|\beta\|_2^2 + \frac{C}{2} \sum_{i=1}^N \|\xi_i\|^2 - \sum_{i=1}^N (h(X_i)\beta - y_i^T + \xi_i^T) \alpha_i \quad (9-4)$$

در رابطه (۹-۴)،  $\alpha_i = [\alpha_1, \dots, \alpha_L]$  پارامترهای نرون‌های لایه پنهان می‌باشد. ماتریس وزن‌های خروجی  $\beta$  با مشتق‌گیری از این رابطه نسبت به  $\alpha_i$ ،  $\xi_i$  و  $\beta$  و مساوی صفر قرار دادن بدست می‌آید. نسخه عادی و هسته بدست آمده برای  $\beta$  به ترتیب در روابط (۱۰-۴) و (۱۱-۴) نشان داده شده است.<sup>۴</sup>

$$\beta = \left( \frac{I}{C} + H^T H \right)^{-1} H^T Y \quad (10-4)$$

<sup>۱</sup> Gradient descend

<sup>۲</sup> Tuning free

<sup>۳</sup> Regularization factor

<sup>۴</sup> Karush-Kuhn-Tucker (KKT)

<sup>۵</sup> جزئیات بیشتر در مورد نحوه محاسبات صورت گرفته و بدست آمدن هر رابطه در [۵۲].

$$\beta = H^T \left( \frac{I}{C} + H^T H \right)^{-1} Y = H^T \left( \frac{I}{C} + KM \right)^{-1} Y \quad (11-4)$$

در روابط (10-۴) و (11-۴)،  $I$  ماتریس همانی و

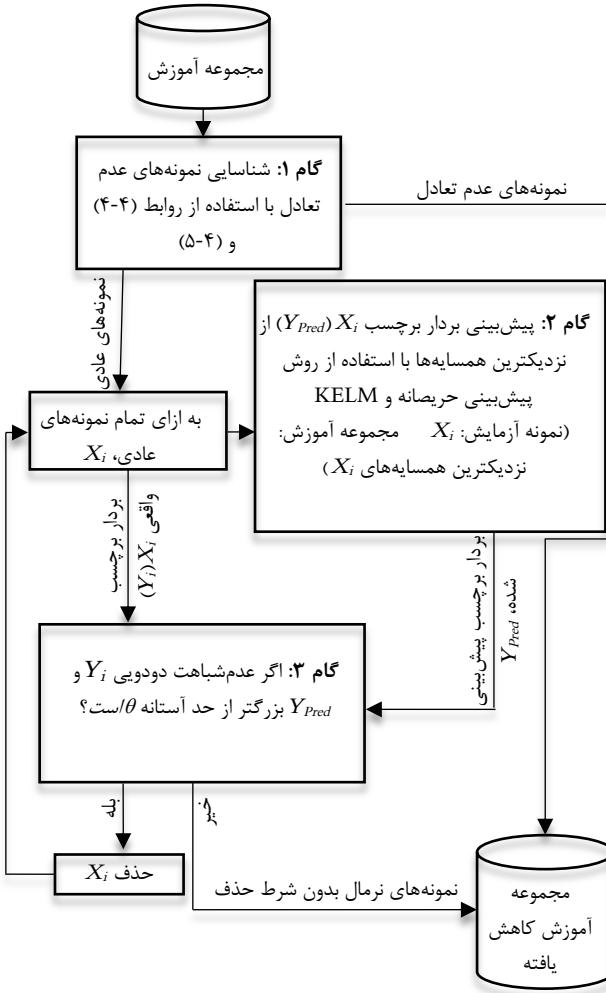
$KM = \{KM(X_i, X_j)\} = \{h(X_i).h(X_j)\}$ ،  $i, j = 1 \dots N$  ماتریس هسته نمونه‌های آموزش می‌باشد. پس از بدست آمدن ماتریس وزن‌های خروجی  $\beta$  در مرحله آموزش، رتبه هر کلاس به ازای هر نمونه آزمایش با استفاده از رابطه (6-۴) بدست می‌آید. سپس در مورد نحوه برچسب‌گذاری تصمیم‌گیری می‌شود. در مسائل تک کلاسه بزرگترین رتبه و در مسائل چند کلاسه و چندبرچسبه بر اساس یک آستانه خاص فرآیند برچسب‌گذاری صورت می‌گیرد. (نحوه برچسب‌گذاری در قسمت ۲-۳-۴ مورد بحث قرار خواهد گرفت)

صلاحیت بکارگیری نسخه هسته روش ELM در مقایسه با سایر الگوریتم‌های یادگیری از جمله نسخه‌های متعدد SVM (مانند LS-SVM و P-SVM) و شبکه‌های عصبی تک لایه در [۵۲] نشان داده شده است. همچنین بر اساس نتایج آزمایش‌های ارائه شده، KELM دارای دقت طبقه‌بندی بهتری نسبت به ELM می‌باشد. از این‌رو در مرحله پیش‌بینی برچسب از روی نزدیکرین همسایه‌ها در روش PS ارائه شده و در مرحله شرح‌گذاری تصاویر بدون برچسب از این روش استفاده شده است.

#### ۴-۲-۲-۴ روش پیشنهادی MLENN-GKELM

در این بخش، روش پیشنهادی دوم در انتخاب نمونه‌های اولیه چندبرچسبه ارائه می‌شود. روش PS چندبرچسبه ENN مبتنی بر پیش‌بینی حریصانه و <sup>۱</sup>KELM شامل ۳ گام اصلی می‌باشد. اساس کار این روش مبتنی بر ENN با تغییر در مرحله تصمیم‌گیری در مورد حذف نمونه‌های کاندیدا با ایده‌های مطرح شده است. شکل ۳-۴ شمای کلی فرآیند انتخاب نمونه‌های اولیه توسط رویکرد MLENN-GKELM ارائه شده را نشان می‌دهد.

<sup>۱</sup> Multi Label Edited Nearest Neighbor based on Greedy and KELM prediction (MLENN-GKELM)



شکل ۳-۴: شماتی کلی گام‌های روش MLENN-GKELM ارائه شده.

همانطور که شکل ۳-۴ نشان می‌دهد، سه گام این روش شامل شناسایی و نگهداری نمونه‌های عدم تعادل، پیش‌بینی برچسب سایر نمونه‌ها از نزدیکترین همسایه‌ها و مقایسه بردار برچسب پیش‌بینی شده و واقعی با روابط عدم شباهت دودویی هستند. در ادامه سه گام اصلی برچسب پیش‌بینی شده و واقعی با روابط عدم شباهت دودویی هستند. در ادامه سه گام اصلی مطرح شده را تشریح خواهیم کرد.

### گام ۱: شناسایی و نگهداری نمونه‌های منجر به عدم تعادل مجموعه آموزش:

مجموعه آموزش مقیاس بزرگ شامل تعداد بالای نمونه‌های چندبرچسبی به الگوریتم MLENN-GKELM داده می‌شود. در گام اول، نمونه‌های آموزش عدم تعادل شناسایی می‌شوند. همانطور که بیان شد، در هر مجموعه چندبرچسبی با افزایش تعداد نمونه‌ها و افزایش تعداد برچسب‌ها سطح عدم تعادل بالا می‌رود [۱۲]. در این مجموعه‌ها تعدادی از نمونه‌ها منجر به عدم تعادل شده و

عدم وجود این نمونه‌ها سطح عدم تعادل مجموعه را بالا می‌برد. در نتیجه دقت بکارگیری هر روش طبقه‌بندی چندبرچسبه را کاهش می‌دهند (جزئیات در بخش ۲-۱-۴- بیان شده است). از این‌رو باستی این نمونه‌ها شناسایی شده و به مجموعه کاهش یافته اضافه شوند. در این گام با استفاده از ایده مطرح شده در [۱۲] این نمونه‌ها شناسایی شده و به زیرمجموعه کاهش یافته اضافه می‌شوند. شناسایی این نمونه‌ها بر اساس روابط (۴-۴) و (۵-۴) انجام می‌شود. به عبارت دیگر زیرمجموعه کاهش یافته پیش از انجام هرگونه عملیاتی شامل این نمونه‌ها خواهد بود.

### گام ۲: پیش‌بینی برچسب نمونه‌های عادی از K نزدیکترین همسایه‌های آن‌ها: پس

از شناسایی نمونه‌های عدم تعادل در گام قبل، تصمیم‌گیری در مورد حضور و یا عدم حضور نمونه‌های باقیمانده در زیرمجموعه کاهش یافته صورت می‌گیرد. بدین ترتیب در ابتدا به ازای هر نمونه آموزش باقیمانده، K نزدیکترین همسایه‌های آن در بین همه نمونه‌های آموزش یافت می‌شود. سپس بردار برچسب نمونه کاندیدای حذف ساخته می‌شود. ایده اصلی این گام در روش MLENN-GKELM پیشنهادی، بکارگیری روش پیش‌بینی حریصانه و KELM در پیش‌بینی برچسب نمونه کاندیدای حذف از K نزدیکترین همسایه‌هایش می‌باشد. اساس کار بکارگیری این دو روش در این گام، فرض پیش‌بینی دقیق بردار برچسب نمونه‌ها می‌باشد. بدین ترتیب این دو روش پیش‌بینی با پارامترهای تنظیم شده یادگیری شده و بردار برچسب نمونه کاندیدای حذف پیش‌بینی می‌شود. پیش‌بینی چندبرچسبه بکار گرفته شده در این بخش بر اساس روابط آستانه‌گذاری مطرح شده در بخش ۳-۲ می‌باشد.

### گام ۳: مقایسه بردار برچسب پیش‌بینی شده و بردار برچسب واقعی: همانطور که

قبل بیان شد، اساس کار روش انتخاب نمونه اولیه پایه‌ای استفاده شده در MLENN-GKELM در تصمیم‌گیری حضور و یا عدم حضور یک نمونه مبتنی بر قاعده نزدیکترین همسایه‌ها می‌باشد. در این تصمیم‌گیری برچسب نمونه با برچسب پیش‌بینی شده از نزدیکترین همسایه‌هایش مقایسه شده و در صورت عدم شباهت، نمونه از مجموعه حذف می‌شود [۱۲] و [۳۶]. در مرحله تصمیم‌گیری حضور و یا عدم حضور یک نمونه در روش MLENN-GKELM ارائه شده، اثربخشی روابط متعدد محاسبه عدم شباهت دودویی ارائه شده در جدول ۱-۴ مورد بررسی قرار گرفته است.

این روابط بر اساس کیفیت محاسبه عدم شباهت مورد مقایسه قرار گرفته‌اند. برخی از این روابط مانند فاصله همینگ دارای منطق محاسبه ساده ولی دقت محاسبه پایین می‌باشند. برخی دیگر به مانند فاصله وتری دارای منطق پیچیده‌تر و دقت بالاتری هستند. در میان روابط بررسی شده، بهترین رابطه از لحاظ دقت محاسبه انتخاب شده است. در مورد نحوه انتخاب بهترین رابطه در بخش نتایج تجربی بحث شده است.

در نهایت، خروجی این الگوریتم، زیرمجموعه کاهش یافته شامل نمونه‌های اولیه است. در فرآیند طبقه‌بندی نمونه‌های بدون برچسب از این زیرمجموعه به عنوان مجموعه آموزش استفاده خواهد شد. الگوریتم ۲-۴ شبه کد روش MLENN-GKELM ارائه شده را نشان می‌دهد. ورودی این الگوریتم مجموعه آموزش و خروجی آن زیرمجموعه آموزش کاهش یافته شامل نمونه‌های اولیه می‌باشد. همچنین  $K$  و  $\theta_1$  دو پارامتر ورودی این الگوریتم هستند.  $K$  تعداد نزدیکترین همسایه‌ها است. از طرفی  $\theta_1$  نشان دهنده حد آستانه تصمیم‌گیری درمورد حذف نمونه‌ها است. خروجی  $k$ ، به عنوان مجموعه آموزش در طبقه‌بندی نمونه‌های بدون برچسب در نظر گرفته می‌شود. در مورد نحوه طبقه‌بندی در بخش بعد بحث خواهیم کرد.

---

### MLENN-GKELM algorithm

---

**Inputs:**

$X_{train} = [l_1, l_2, l_3, \dots, l_n]$ , Training instances  
 $Y = [Y_1, Y_2, Y_3, \dots, Y_n]$ , Training labels  
 $K$ , Number of nearest neighbor  
 $\Theta_I$ , Threshold value for decision about removing

**Output:**

$S \subseteq X_{train}$

**Process:**

```
Compute LblIR for each label in  $X_{train}$  from equation (4-4)
Compute DSIR for  $X_{train}$  from equation (4-5)
foreach (Instance  $X_i$  in  $X_{train}$ ) do
    if ( $X_i$  has a label with  $LblIR > DSIR$ )
        InstanceMarkRemoving ← 0
    else
        find  $K$  nearest neighbor of  $X_i$ 
         $Y_{pred}$ =Predict label of  $X_i$  from nearest neighbor by Greedy and KELM algorithms
        if (Multi-label binary dissimilarity of  $Y_i$  and  $Y_{pred} > \Theta_I$ )
            InstanceMarkRemoving ← 1
        end
    end
end
S=All instances with InstanceMarkRemoving=0
Return S
```

---

الگوریتم ۴-۲: شبکه کد روش MLENN-GKELM ارائه شده.

## ۴-۲-۵- جمع‌بندی

در این بخش دو رویکرد جهت انتخاب نمونه‌های اولیه چندبرچسبه ارائه شد. این دو رویکرد با تغییر در فرضیه اصلی روش‌های پایه PS، سعی در بهبود عملکرد آن‌ها دارند. از طرفی با بکارگیری ایده‌های مطرح شده در این دو رویکرد، چالش زمان و مقیاس‌پذیری سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ مرتفع می‌گردد.

## ۴-۳- شرح گذاری

پس از کاهش حجم مجموعه آموزش و انتخاب نمونه‌های اولیه در مرحله قبل، شرح‌گذاری تصاویر بدون برچسب صورت می‌گیرد. در مرحله شرح‌گذاری، نمونه‌های بدون برچسب و نمونه‌های اولیه برچسبدار به الگوریتم شرح‌گذاری داده شده و روند شرح‌گذاری صورت می‌گیرد. شرح‌گذاری تصاویر در این پژوهش با سه الگوریتم صورت گرفته است. روش KNN، روش پایه‌ای شرح‌گذاری در این پایان‌نامه می‌باشد [۵]. به منظور بهبود دقت KNN، از روش‌های متعدد وزن‌دهی استفاده

شده است [۶۴-۶۵]. از این‌رو برای بهبود دقت شرح‌گذاری در این پایان‌نامه نیز این روش‌ها بکار گرفته شده است. در نهایت از روش یادگیری KELM نیز در جهت بهبود دقت شرح‌گذاری تصاویر بدون برچسب استفاده شده است. در قسمت بعد در ابتدا مختصری در مورد این روش‌ها توضیح خواهیم داد. سپس رویکردهای آستانه‌گذاری بکار رفته در این پایان‌نامه در تخصیص برچسب چندگانه به نمونه‌ها را بررسی خواهیم کرد.

#### ۱-۳-۴ شرح گذاری مبتنی بر K-نزدیکترین همسایه وزن دار

الگوریتم KNN یکی از روش‌های ساده و قدرتمند در حوزه یادگیری ماشین و تشخیص الگو می‌باشد [۶۵]. بکارگیری این روش در کاربردهای تجزیه تحلیل و فهم محتوای تصاویر منجر به بهبود دقت طبقه‌بندی نسبت به دیگر روش‌های یادگیری می‌شود [۲]. ایده کلی این روش اختصاص برچسب‌های مفهومی مشابه به نمونه‌های مشابه در فضای ویژگی است. در این روش بر اساس معیار فاصله مشخص،  $K$  نزدیکترین همسایه نمونه بدون برچسب  $u_i$ ، محاسبه می‌شود. سپس با رای‌گیری حداکثری از این همسایه‌ها، برچسب‌های نمونه  $u_i$  بدست می‌آید.

در رای‌گیری حداکثری در نسخه عادی روش KNN، تاثیر حضور و رای‌دهی همسایه‌های اول تا  $K$ ام برابر فرض می‌شود. در صورتی که با افزایش تعداد نزدیکترین همسایه‌ها، در نظر گرفتن تاثیر برابر همسایه اول تا  $K$ ام منجر به افت کارایی روش می‌شود. به منظور غلبه بر این مشکل، نسخه وزن دار روش KNN ارائه شده است [۶۴]. در KNN وزن دار به  $K$  نزدیکترین همسایه نمونه  $u_i$  وزن و رتبه‌های مختلفی اختصاص می‌یابد. مقدار وزن اختصاص داده شده با افزایش فاصله مابین نمونه  $u_i$  و نمونه‌های همسایه کاهش می‌یابد. به عبارت دقیق‌تر به نزدیکترین همسایه بیشترین وزن و به  $K$ امین همسایه کمترین وزن اختصاص می‌یابد. تاکنون روابط وزن‌دهی مختلفی در روش KNN وزن دار ارائه شده است [۶۲]. از آنجا که یکی از روش‌های شرح‌گذاری بکاررفته در این پایان‌نامه نسخه وزن دار روش KNN است، در این بخش روابط وزن‌دهی مختلف در این روش مورد بررسی قرار می‌گیرد. در این روابط فاصله نمونه بدون برچسب  $u_i$  با  $K$  نزدیکترین

همسايه‌هايش بصورت  $d_{i,j}$  می‌باشند. همچنین  $d_{i,1}, d_{i,2}, d_{i,3}, \dots, d_{i,K}$  فاصله نمونه بدون برچسب  $u_i$  از همسایه زام و  $w_j$  وزن این همسایه می‌باشد.

- **وزن دهی خطی<sup>۱</sup>:** اولین و معروفترین روش وزن دهی توسط دودانی<sup>۲</sup> ارائه شده است [۶۶].

این روش، وزن دهی خطی با در نظر گرفتن اهمیت بیشتر برای نمونه‌های نزدیکتر می‌باشد. وزن همسایه زام نمونه بدون برچسب  $u_i$  در این روش با استفاده از رابطه زیر بدست می‌آید.

$$w_j = \begin{cases} \frac{d_{i,k} - d_{i,j}}{d_{i,k} - d_{i,1}} & \text{if } d_{i,k} = d_{i,1} \\ 1 & \text{if } d_{i,k} \neq d_{i,1} \end{cases} \quad (12-4)$$

- **وزن دهی خطی دوگان<sup>۳</sup>:** این روش وزن دهی در [۶۳] ارائه شده است و رابطه وزن دهی تعریف شده، دوگان روش قبلی است. وزن همسایه زام نمونه بدون برچسب  $u_i$  در این روش با استفاده از رابطه زیر بدست می‌آید.

$$w_j = \begin{cases} \frac{d_{i,k} - d_{i,j}}{d_{i,k} - d_{i,1}} * \frac{d_{i,k} + d_{i,1}}{d_{i,k} + d_{i,j}} & \text{if } d_{i,k} = d_{i,1} \\ 1 & \text{if } d_{i,k} \neq d_{i,1} \end{cases} \quad (13-4)$$

- **وزن دهی خطی معکوس<sup>۴</sup>:** در این روش، وزن هر نمونه در مجموعه نزدیکترین همسایه‌ها، معکوس فاصله آن نمونه تا نمونه بدون برچسب می‌باشد. بنابراین وزن همسایه زام نمونه بدون برچسب  $u_i$  با استفاده از رابطه زیر بدست می‌آید [۶۲].

$$w_j = \frac{1}{d_{i,j}} \quad \text{if } d_{i,j} \neq 0 \quad (14-4)$$

- **وزن دهی مرتبه<sup>۵</sup>:** نحوه وزن دهی در این روش بر اساس رتبه و جایگاه هر نمونه در مجموعه نزدیکترین همسایه‌ها می‌باشد. وزن همسایه زام نمونه بدون برچسب  $u_i$  در این روش با استفاده از رابطه زیر بدست می‌آید [۶۲].

<sup>1</sup> Linear weighting

<sup>2</sup> Dudani

<sup>3</sup> Dual linear weighting

<sup>4</sup> Inverse linear weighting

<sup>5</sup> Rank weighting

$$w_j = K - j + 1 \quad (15-4)$$

در این رابطه  $K$  تعداد نزدیکترین همسایه‌ها و ز جایگاه نمونه همسایه زام می‌باشد.

- وزن دهی لگاریتمی<sup>۱</sup>: نحوه وزن دهی در این روش بر اساس معکوس لگاریتم فاصله مابین نمونه بدون برچسب و نمونه‌های مجموعه نزدیکترین همسایه می‌باشد. بنابراین در این روش وزن همسایه زام نمونه بدون برچسب  $X_i$  با استفاده از رابطه زیر بدست می‌آید.

$$w_j = \frac{1}{\log((1 + \tau * j) * d_{i,j})} \quad (16-4)$$

در این رابطه پارامتر  $\tau$  به منظور کنترل تاثیر رتبه همسایه زام در مجموعه نزدیکترین همسایه‌ها بوده و به صورت تجربی مقداردهی می‌شود [۲].

پس از محاسبه وزن هر نمونه همسایه در مجموعه نزدیکترین همسایه‌ها با استفاده از روابط مطرح شده، وزن‌های بدست آمده در بردار برچسب هر نمونه همسایه ضرب شده و بردار رتبه هر نمونه همسایه بدست می‌آید. به عبارت دیگر هر نمونه همسایه دارای یک برچسب مشخص، فقط به اندازه وزن خود در رای دهی به برچسب مورد نظر مشارکت می‌کند. در مرحله بعد همه بردار رتبه‌های حاصل از هر نمونه در مجموعه نزدیکرین همسایه‌ها با یکدیگر جمع شده و بردار واحدی که نشان دهنده رتبه نمونه بدون برچسب  $w_i$  است، تشکیل می‌شود. از این‌رو خروجی الگوریتم KNN وزن‌دار بردار رتبه نمونه بدون برچسب  $w_i$  می‌باشد که بایستی تصمیم‌گیری در مورد اختصاص برچسب بر اساس رتبه‌های بدست آمده بر روی آن صورت گیرد. در مورد تصمیم‌گیری چندبرچسبه در بخش ۳-۴-۳-۴ بحث خواهیم کرد.

#### ۲-۳-۴ شرح‌گذاری مبتنی بر الگوریتم KELM

فرض اصلی بکارگیری روش KELM در مرحله شرح‌گذاری در این پایان‌نامه، صلاحیت عملکرد آن نسبت به سایر نسخه‌های ارائه شده از ELM و الگوریتم‌های یادگیری می‌باشد [۵۲]. به منظور طبقه‌بندی نمونه‌های بدون برچسب، در ابتداء الگوریتم KELM (توصیف شده در بخش ۴-۴ منظور

---

<sup>۱</sup> Logarithmic weighting

۳-۲) با استفاده از مجموعه آموزش کاهش یافته و پارامترهای تنظیم شده یادگیری شده و ماتریس وزن‌های خروجی بدست می‌آید. سپس مقدار رتبه نمونه‌های بدون برچسب، با ضرب ماتریس فاصله هسته آنها در ماتریس ضرایب  $\beta$  بدست می‌آید. در نهایت برچسب‌گذاری نمونه‌های بدون برچسب بر اساس رتبه‌های بدست آمده صورت می‌گیرد. در این الگوریتم تنظیم پارامترهای مناسب بر اساس ویژگی‌های هر مجموعه داده از اهمیت بالای برخوردار است. در فصل بعد در مورد نحوه تنظیم پارامترهای مرتبط با هر بخش توضیح خواهیم داد.

### ۳-۳-۴ آستانه‌گذاری در طبقه‌بندی چندبرچسبی

پس از ساخته شدن بردار رتبه نهایی توسط الگوریتم‌های یادگیری مرحله قبل، تصمیم‌گیری در مورد اختصاص یک برچسب به نمونه بدون برچسب  $X_i$  صورت می‌گیرد. روش‌های متعددی به منظور تصمیم‌گیری چندبرچسبی ارائه شده است [۶۷]. این روش‌ها بر اساس رویکردهای آستانه‌گذاری روند برچسب‌گذاری را انجام می‌دهند. در نظر گرفتن مقدار آستانه ثابت منجر به افت کارایی و عملکرد روش‌های طبقه‌بندی چندبرچسبی می‌شود. در فرآیند شرح‌گذاری در این پایان‌نامه دو رویکرد آستانه‌گذاری در تصمیم‌گیری چندبرچسبی بکار رفته است. رویکرد اول، آستانه‌گذاری درصد جز به کل و رویکرد دوم، روشی تطبیقی است. فرض کنید بردار رتبه بدست آمده در مرحله قبل به صورت  $SC_i = [score_1, score_2, score_3, \dots score_c]$  باشد. این بردار شامل  $c$  مقدار رتبه متناظر با  $c$  برچسب می‌باشد. رویکرد آستانه‌گذاری درصد جز به کل بکار رفته به منظور ساخت بردار برچسب باینتری نمونه  $X_i$  از روی بردار رتبه،  $m$  درصد از بزرگترین رتبه ها را بر اساس رابطه (۱۷-۴) برابر یک قرار داده و مابقی را صفر در نظر می‌گیرد.

$$\frac{score_1 + score_2 + score_3 + \dots + score_l}{score_1 + score_2 + score_3 + \dots + score_c} \geq \theta_2 \quad (17-4)$$

به بیان دیگر این رویکرد  $l$  مقدار از  $c$  مقدار رتبه‌هایی که  $\theta_2$  درصد از کل مقادیر را شامل می‌شوند، برابر یک قرار می‌دهد.

دومین رویکرد تصمیم‌گیری چندبرچسبه بکار رفته در این بخش، رویکرد آستانه‌گذاری تطبیقی است. در این رویکرد به منظور ساخت بردار برچسب دودویی نمونه  $X_i$ ، رتبه‌هایی را که  $\theta_3$  درصد مجموع همه رتبه‌ها باشند برابر یک قرار داده و مابقی را صفر در نظر می‌گیرد. رابطه

(۱۸-۴) فرآیند آستانه‌گذاری تطبیقی مورد استفاده را نشان می‌دهد.

$$\begin{cases} Y_i(j) = 1 & \text{if } SC_i(j) \geq \theta_3 * \text{sum}(SC_i) \\ Y_i(j) = 0 & \text{Otherwise} \end{cases} \quad (18-4)$$

مقادیر  $\theta_2$  و  $\theta_3$  به صورت تجربی و بر اساس ویژگی‌های هر مجموعه تعیین می‌شوند. از طرفی این مقادیر رابطه مستقیمی با میانگین برچسب‌های هر نمونه در مجموعه داده دارند. به عبارت دیگر هر چه میانگین تعداد برچسب‌های هر نمونه درون مجموعه بیشتر باشد، مقدار بیشتری برای  $\theta_2$  و  $\theta_3$  نیز بایستی در نظر گرفته شود. با تنظیم درست این مقادیر، بردار برچسب قابل اعتمادی برای هر نمونه بدون برچسب بدست خواهد آمد. در بخش نتایج تجربی به نحوه تنظیم  $\theta_2$  و  $\theta_3$  با توجه به خصوصیات هر مجموعه پرداخته خواهد شد.

#### ۴-۳-۴ جمع‌بندی

در این بخش دو روش شرح‌گذاری و رویکردهای آستانه‌گذاری تصمیم‌گیری چندبرچسبه بکار رفته در این پایان‌نامه تشریح شد. در فرآیند شرح‌گذاری تصاویر بدون برچسب از این روش‌ها و رویکردهای آستانه‌گذاری معرفی شده استفاده می‌شود. مرجع آموزش شرح‌گذاری در این روش‌ها، مجموعه آموزش کاهش یافته شامل نمونه‌های اولیه انتخاب شده می‌باشد.

#### ۴-۴- جمع‌بندی رویکردهای ارائه شده

فرآیند کلی شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ در این پایان‌نامه در دو مرحله عمده مورد بررسی قرار گرفت. در مرحله انتخاب نمونه‌های اولیه، مجموعه آموزش پیش‌پردازش می‌شود. پیش‌پردازش صورت گرفته شامل کاهش مجموعه آموزش و انتخاب نمونه‌های موثر در فرآیند یادگیری با دو روش COCNN و MLENN-GKELM ارائه شده می‌باشد. خروجی فرآیند پیش‌پردازشی به عنوان مرجع روش‌های یادگیری در مرحله شرح‌گذاری در نظر گرفته می‌شود. به

منظور بهبود دقت شرحگذاری در این پژوهش نیز از دو روش یادگیری KNN وزن دار و KELM استفاده شد. همچنین از دو رویکرد آستانه گذاری در تصمیمگیری چندبرچسبه بهره برده شده است. در ادامه این پایان نامه و در فصل بعد به نحوه پیاده سازی و بیان نتایج ارزیابی ایده های ارائه شده خواهیم پرداخت.

## فصل ۵:

### نتایج تجربی

## ۱-۵ - پیشگفتار

در این فصل به بیان نتایج پیاده‌سازی رویکردهای پیشنهادی می‌پردازیم. بدین منظور در ابتدا مجموعه تصاویر و معیارهای ارزیابی مورد استفاده را معرفی خواهیم کرد. سپس به بیان تنظیم‌های اولیه آزمایش‌ها و ارائه نتایج خواهیم پرداخت.

## ۲-۵ - مجموعه تصاویر مورد استفاده

به منظور ارزیابی رویکردهای ارائه شده در این پایان‌نامه در حل چالش مقیاس‌پذیری و دقت شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ، از مجموعه تصویر مقیاس بزرگ NUS-WIDE و نسخه‌های متعدد آن استفاده شده است [۵]. در ادامه این مجموعه تصویر و نسخه‌های آنرا توصیف خواهیم کرد.

**NUS-WIDE:** این مجموعه شامل ۲۶۹۶۴۸ تصویر جمع‌آوری‌شده از سایت <sup>۱</sup>Flickr به همراه تگ‌های موجود برای هر تصویر در این سایت می‌باشد. به منظور حذف اختلال تگ‌ها، ارائه‌کنندگان این مجموعه ۸۱ مفهوم از تگ‌ها را استخراج کرده و تصاویر مربوطه را به صورت نیمه‌خودکار با این ۸۱ مفهوم شرح‌گذاری کرده و در اختیار پژوهشگران حوزه شرح‌گذاری تصاویر در مقیاس بزرگ قرار داده‌اند. از ۲۶۹۶۴۸ تصویر موجود در این مجموعه، ۱۶۱۷۸۹ تصویر به عنوان مجموعه آموزش و مابقی یعنی ۱۰۷۸۵۹ تصویر به عنوان مجموعه آزمایش انتخاب شده است.

**NUS-WIDE-Object:** نسخه NUS-WIDE-Object شامل ۳۰۰۰۰ تصویر در ۳۱ طبقه از اشیا می‌باشد. از ۳۰۰۰۰ تصویر موجود در این نسخه، ۱۷۹۲۸ تصویر به عنوان آموزش و ۱۲۰۷۲ تصویر به عنوان آزمایش انتخاب شده است.

**NUS-WIDE-Scene:** نسخه NUS-WIDE-Scene شامل ۳۴۹۲۶ تصویر در ۳۳ مفهوم صحنه می‌باشد. از ۳۴۹۲۶ تصویر موجود در این نسخه، ۱۷۴۶۳ تصویر به عنوان آموزش و ۱۷۴۶۳ تصویر به عنوان آزمایش انتخاب شده است.

<sup>1</sup> [www.flickr.com](http://www.flickr.com)

## ۵۵۶۱۵: درنهایت نسخه NUS-WIDE-Lite زیرمجموعه‌ای شامل

تصویر انتخاب شده از کل ۸۱ مفهوم موجود در نسخه اصلی است. از ۵۵۶۱۵ تصویر موجود در این نسخه، ۲۷۸۰۷ تصویر به عنوان آموزش و ۲۷۸۰۸ تصویر به عنوان آزمایش انتخاب شده است.

### ۳-۵- توصیف تصاویر

به منظور انجام فرآیند شرح‌گذاری ابتدا لازم است تا بردار ویژگی دیداری سطح پایین تصاویر ساخته شود. روش‌های مختلفی برای توصیف تصاویر و ساخت بردار ویژگی وجود دارد. در یک دسته‌بندی کلی می‌توان آن‌ها را به دو دسته سراسری<sup>۱</sup> و محلی<sup>۲</sup> تقسیم کرد. در ادامه به‌طور مختصر در مورد هر یک بحث خواهیم کرد.

**ویژگی‌های سراسری:** ویژگی‌های سراسری، ویژگی‌هایی است که از کل تصویر استخراج می‌شوند. به عبارت دیگر محاسبه بردار ویژگی سراسری با اعمال روش استخراج ویژگی روی کل تصویر ورودی انجام می‌شود. ویژگی‌های سراسری رایج که از تصاویر استخراج می‌شود شامل رنگ، بافت، شکل و یا ویژگی‌های ترکیبی هستند [۶۸].

**ویژگی‌های محلی:** برای محاسبه ویژگی‌های محلی برعکس ویژگی‌های سراسری، تصویر را ناحیه<sup>۳</sup> بندی کرده و عملیات استخراج ویژگی را از نواحی تصویر انجام می‌دهند. به این نواحی درون تصویر توصیفگرهای محلی<sup>۴</sup> گفته می‌شود [۶۸]. از پر کاربردترین ویژگی‌های محلی می‌توان به کیسه لغات دیداری<sup>۵</sup> (BoW) اشاره کرد.

در مجموعه تصاویر مورد استفاده در این پژوهش به مانند [۴] و [۱۸] به منظور نمایش هر تصویر از سه دسته ویژگی سراسری معرفی شده در [۵] استفاده می‌شود. این سه دسته ویژگی عبارتند از: (۱) کورولوگرام<sup>۶</sup> رنگ HSV ۱۴۴ بعدی، (۲) هیستوگرام جهت لبه ۷۳ بعدی و (۳۵)

<sup>1</sup> Global

<sup>2</sup> Local

<sup>3</sup> Region

<sup>4</sup> Local descriptors

<sup>5</sup> Bag of Visual Words(BoW)

<sup>6</sup> Correlogram

موجک بافت<sup>۱</sup> بعدی. بردار ویژگی هر تصویر از کنار هم قرار دادن این سه دسته ویژگی بدست آمده است. از این‌رو هر تصویر با یک بردار ۳۴۵ بعدی نمایش داده می‌شود.

#### ۴-۵ معیارهای ارزیابی

به منظور ارزیابی دقت شرح‌گذاری تصاویر و همچنین ارزیابی قدرت عملکرد روش‌های PS معیارهای متفاوتی وجود دارد. دقت شرح‌گذاری، قدرت کاهش حجم نمونه‌ها و همچنین تعادل مابین دقت شرح‌گذاری و کاهش حجم نمونه‌ها از جمله معیارهای ارزیابی مورد استفاده در این پژوهش هستند. در این بخش به معرفی معیارهای ارزیابی مورد استفاده در این پژوهش خواهیم پرداخت.

#### ۱-۴-۵ ارزیابی نتایج شرح‌گذاری

از جمله معیارهای معروف مورد استفاده در ارزیابی مسائل طبقه‌بندی، دقت، فراخوان<sup>۲</sup> و F1 می‌باشد. اگر  $TP$  تعداد نمونه‌های مثبت درست مثبت پیش‌بینی شده،  $TN$  تعداد نمونه‌های منفی درست منفی پیش‌بینی شده،  $FP$  تعداد نمونه‌های منفی به اشتباه مثبت پیش‌بینی شده و  $FN$  تعداد نمونه‌های مثبت به اشتباه منفی پیش‌بینی شده باشد، مقادیر دقت، فراخوان و F1 با استفاده از رابطه‌های (۱-۵)، (۲-۵) و (۳-۵) بدست می‌آیند.

$$Precision = \frac{TP}{TP + FP} \quad (1-5)$$

$$Recall = \frac{TP}{TP + FN} \quad (2-5)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3-5)$$

---

<sup>1</sup> Wavelet Texture

<sup>2</sup> Recall

در مسائل چندبرچسبه به ازای همه برچسبها این معیارها محاسبه شده و میانگین آنها به عنوان معیار ارزیابی نهایی در نظر گرفته می‌شود. به عنوان مثال، میانگین دقت تمام برچسبها<sup>۱</sup> (MAP) محاسبه به عنوان معیار ارزیابی در نظر گرفته می‌شود.

در میانگین‌گیری معیارهای ارزیابی در مسائل چندبرچسبه از دو رویکرد آماری میانگین‌گیری میکرو<sup>۲</sup> و ماکرو<sup>۳</sup> استفاده می‌شود. در میانگین‌گیری ماکرو، ابتدا بر اساس مقادیر  $FP$ ,  $TN$ ,  $TP$  و  $FN$  معیار ارزیابی مورد نظر هر برچسب به صورت جداگانه محاسبه شده و سپس از مقادیر بدست آمده میانگین‌گیری می‌شود. در صورتی که در میانگین‌گیری میکرو در ابتدا از مقادیر  $FP$ ,  $TN$ ,  $TP$  و  $FN$  میانگین‌گیری شده و سپس معیار ارزیابی محاسبه می‌شود [۶۹]. رابطه‌های (۴-۵) و (۵-۵) نحوه محاسبه معیارهای ارزیابی میکرو و ماکرو را نشان می‌دهند. در این رابطه‌ها  $c$  نشان دهنده تعداد برچسب‌هاست.

$$measure_{macro} = \frac{\sum_{i=1}^c measure(TP_i, TN_i, FP_i, FN_i)}{c} \quad (4-5)$$

$$measure_{micro} = measure(\sum_{i=1}^c TP_i, \sum_{i=1}^c TN_i, \sum_{i=1}^c FP_i, \sum_{i=1}^c FN_i) \quad (5-5)$$

در این پژوهش از میانگین‌گیری میکرو و ماکرو استفاده شده است. بدین ترتیب مقادیر دقت ماکرو، فراخوان ماکرو، دقت میکرو، فراخوان میکرو و F1 میکرو محاسبه و به عنوان معیار ارزیابی نتایج شرح‌گذاری استفاده شده است.

## ۲-۴-۵ ارزیابی نتایج انتخاب نمونه‌های اولیه

در بکارگیری روش‌های PS علاوه بر دقت طبقه‌بندی که با استفاده از معیارهای تعریف شده در بالا محاسبه می‌شود، میزان کاهش حجم نمونه‌ها و همچنین کاهش حجم به همراه عدم کاهش

<sup>1</sup> Mean Average Precision (MAP)

<sup>2</sup> Micro averaging

<sup>3</sup> Macro averaging

دقت طبقه‌بندی نیز از اهمیت بالایی برخوردار است. از این‌رو در ارزیابی این روش‌ها از معیارهای ارزیابی دیگری با نام نرخ کاهش<sup>۱</sup> و نرخ تعادل<sup>۲</sup> استفاده می‌شود [۳۱].

نرخ کاهش به صورت تعداد نمونه‌های اولیه بدست آمده توسط هر الگوریتم PS بر تعداد نمونه‌های موجود در مجموعه آموزش اولیه تعریف می‌شود. نرخ تعادل نیز میزان تعادل مابین دقت طبقه‌بندی با بکارگیری مجموعه کاهش یافته به عنوان مجموعه آموزش و نرخ کاهش بدست آمده توسط روش PS را نشان می‌دهد. رابطه‌های (۶-۵) و (۷-۵) به ترتیب نحوه محاسبه این دو معیار را نشان می‌دهند.

$$Reduction\ Rate(RR) = \frac{number\ of\ instance\ after\ apply\ PS\ algorithm}{number\ of\ training\ instance} \quad (6-5)$$

$$Trade-off\ Rate = (Macro - Micro)F1 * RR \quad (7-5)$$

به منظور نشان دادن مقیاس‌پذیری رویکرد پیشنهادی این پژوهش، علاوه بر معیارهای بالا از معیار دیگری بنام زمان شرح‌گذاری هر تصویر استفاده شده است. این زمان بصورت میانگین زمان شرح‌گذاری هر تصویر تعریف می‌شود. هر چقدر این زمان کمتر باشد، مقیاس‌پذیری رویکرد پیشنهادی بالاتر است. آنچه مسلم است هر چه حجم مجموعه آموزش کمتر باشد، زمان شرح‌گذاری هر تصویر نیز کمتر خواهد شد.

ذکر این نکته ضروری است که کلیه پیاده‌سازی‌ها و ارزیابی‌های این پایان‌نامه با استفاده از نرم افزار MATLAB R2013b در سیستم CPU CORE i5 با مشخصه‌های COCNN و MLENN- شرح‌گذاری است. در انتخاب نمونه‌های اولیه پارامترهای بهینه دو رویکرد RAM 6GB صورت گرفته است.

## ۵-۵- تنظیم اولیه آزمایش‌ها

در این قسمت به بیان تنظیم‌های اولیه آزمایش‌های صورت گرفته در این پایان‌نامه می‌پردازیم. این تنظیم‌ها شامل دو بخش تنظیم‌های مربوط به انتخاب نمونه‌های اولیه و شرح‌گذاری است. در انتخاب نمونه‌های اولیه پارامترهای بهینه دو رویکرد COCNN و

<sup>1</sup> Reduction rate

<sup>2</sup> Trade-off rate

GKELM ارائه شده تنظیم می‌شوند. همچنین بررسی اثربخشی بکارگیری روابط محاسبه عدم شباهت دودویی مختلف و اثربخشی بکارگیری توابع هسته مختلف در روش MLENN-GKELM ارائه شده، از دیگر بخش‌های تنظیم‌های مربوط به انتخاب نمونه‌های اولیه هستند. در شرح‌گذاری، اثربخشی بکارگیری روابط وزن‌دهی مختلف در KNN وزن‌دار و همچنین پارامترهای بهینه روش‌های شرح‌گذاری مورد استفاده در این پایان‌نامه مورد بررسی قرار گرفته‌اند. در تمامی آزمایش‌های تنظیم پارامتر و بررسی اثربخشی صورت گرفته، از روش اعتبارسنجی متقابل دو بخشی<sup>۱</sup> بر روی مجموعه آموزش استفاده شده است. در این روش اعتبارسنجی، مجموعه آموزش به دو بخش تقسیم شده و هر بار یکی از بخش‌ها به عنوان مجموعه آموزش و دیگری به عنوان مجموعه اعتبارسنجی در نظر گرفته شده و الگوریتم اجرا می‌شود. سپس مجموعه‌های آموزش و اعتبارسنجی معکوس شده و الگوریتم مجدد اجرا می‌شود. جدول ۱-۵ پارامترهای بهینه تنظیم شده انتخاب نمونه‌های اولیه و شرح‌گذاری را به صورت خلاصه نشان می‌دهد. در ادامه این بخش جزئیات نحوه تنظیم هر پارامتر به تفکیک بیان شده است.

#### ۱-۵-۵ تنظیم‌های انتخاب نمونه‌های اولیه

تنظیم پارامترهای بهینه روش COCNN ارائه شده: در روش COCNN ارائه شده و در محاسبه مقدار  $COF$ ، سه پارامتر  $K$ ,  $\alpha$  و  $\beta$  باید مقداردهی شوند. به منظور تعیین مقدار  $K$  در رابطه (۱-۴)، آزمایشی بر روی مجموعه NUS-WIDE-Object صورت گرفته است. این آزمایش برای سه مقدار مختلف  $K=1$ ,  $K=2$  و  $K=3$  انجام و پارامتر  $K$  مربوط به بهترین مقدار F1 مacro به عنوان پارامتر بهینه انتخاب شده است. در این آزمایش بهترین مقدار F1 Macro مربوط به  $K=2$  بوده و از این‌رو از این مقدار در سایر آزمایش‌های بعدی استفاده شده است.

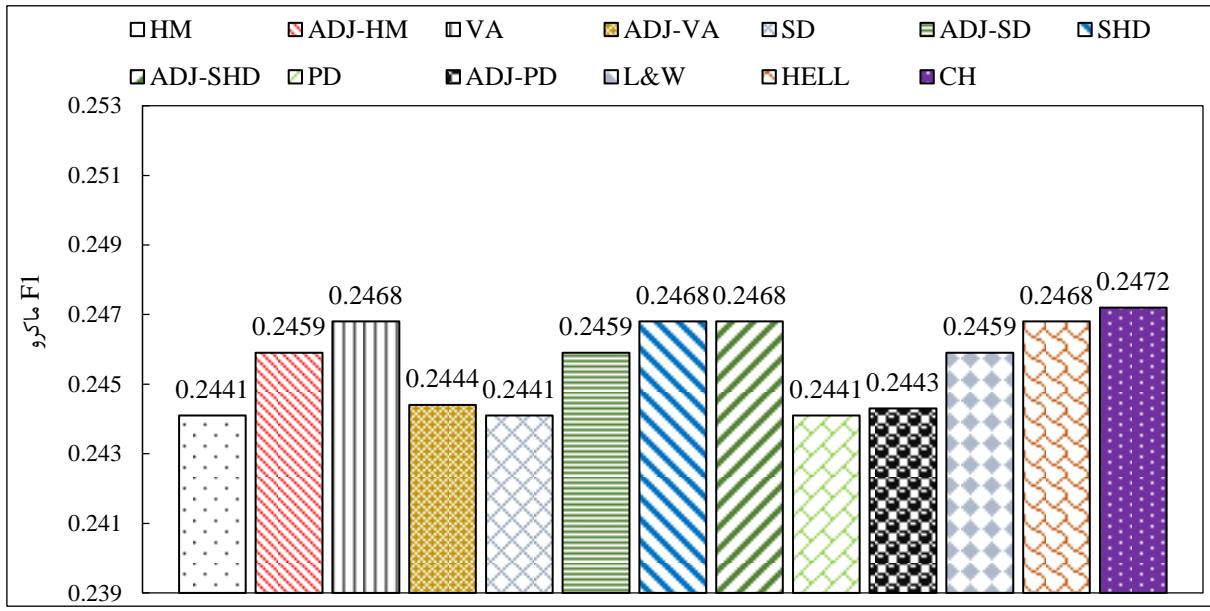
با توجه به استدلال [۴۹] در انتخاب بهینه پارامترهای  $\alpha$  و  $\beta$ ، در صورتی که بیشینه مقدار نمونه‌ها در آزمایش‌های اولیه نزدیک به ۱۰۰ باشد، مقدار بهینه  $\alpha$  نیز برابر ۱۰۰ و همچنین  $Dev$  در صورتی که این مقدار نزدیک به ۱۰ باشد مقدار بهینه  $\alpha$  برابر ۱۰ در نظر گرفته می‌شود. برای

---

<sup>۱</sup> 2-fold cross validation

تعیین مقدار  $\beta$  نیز استدلالی مشابه به کار گرفته می‌شود با این تفاوت که بیشینه مقدار  $KDis$  را در نظر گرفته و مقادیر بهینه این پارامتر برابر  $1/0/0/0/1$  می‌باشد. با توجه به آزمایش‌های اولیه صورت گرفته در این پژوهش و تعیین بیشینه مقدار  $Dev$  و  $KDis$  مقدار  $\alpha$  و  $\beta$  به ترتیب برابر  $100/0/0/1$  در نظر گرفته شده است. مقادیر بهینه تنظیم شده این پارامترها در جدول ۱-۵ نشان داده شده است.

**بررسی اثربخشی بکارگیری روابط عدم شباهت چندبرچسبی متعدد در روش MLENN-GKELM:** به منظور بررسی اثربخشی و تصمیم‌گیری در مورد انتخاب مناسب‌ترین رابطه محاسبه عدم شباهت دودیی در مرحله مقایسه دو بردار برچسب پیش‌بینی شده و بردار NUS-WIDE-برچسب واقعی (گام ۳ روش MLENN-GKELM)، آزمایشی بر روی مجموعه Object صورت گرفته است. در این آزمایش رابطه‌های عدم شباهت دودویی ارائه شده در جدول ۱-۴ مورد مقایسه قرار گرفته‌اند. در ابتدا با استفاده از هر رابطه و با رای‌گیری حداکثری سنتی در گام ۲، (بدون بکارگیری ایده مطرح شده در گام ۲) مجموعه آموزش کاوش یافته و مجموعه آزمایش با استفاده از الگوریتم KNN با  $K=1$  شرح‌گذاری شده است. معیار تصمیم‌گیری در مورد انتخاب مناسب‌ترین تابع محاسبه عدم شباهت دودویی F1 ماکرو در نظر گرفته شده است. شکل ۵-۱ نتایج شرح‌گذاری اثربخشی بکارگیری روابط عدم شباهت دودویی مختلف در روش MLENN-GKELM ارائه شده را نشان می‌دهد.



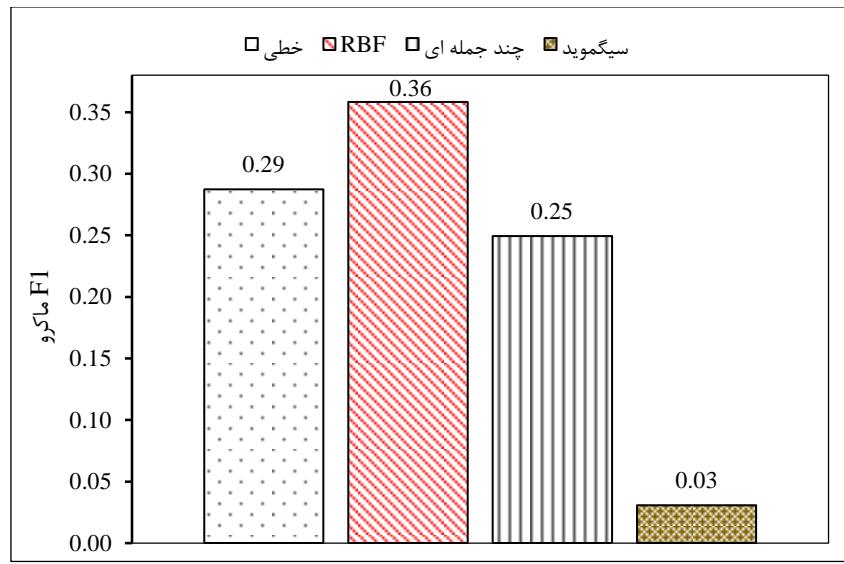
شکل ۱-۵: اثربخشی بکارگیری روابط عدم شباهت دودویی مختلف در روش MLENN-GKELM

همانطور که در شکل ۱-۵ مشاهده می‌شود، رابطه محاسبه عدم شباهت دودویی و تری نسبت به دیگر رابطه‌ها از دقت شرح‌گذاری بالاتری برخوردار است. به عبارت دقیق‌تر، دقت شرح‌گذاری مجموعه آزمایش با استفاده از مجموعه آموزش کاهاش یافته با بکارگیری این رابطه بالاتر از همه رابطه‌های بیان شده است. از این‌رو در سایر آزمایش‌های صورت گرفته در روش MLENN-GKELM پیشنهادی از این رابطه استفاده شده است.

**بررسی اثربخشی بکارگیری توابع هسته مختلف در روش KELM:** روش KELM در دو قسمت از این پژوهش مورد استفاده قرار گرفته است. این روش در پیش‌بینی بردار برچسب نمونه کاندیدای حذف از نزدیکترین همسایه‌ها و همچنین شرح‌گذاری، فرآیند پیش‌بینی را انجام می‌دهد. در این روش پارامتر ضریب کنترل، نوع تابع هسته مورد استفاده و پارامترهای بهینه تابع هسته انتخابی از اهمیت بالایی برخوردار است. در این بخش در ابتدا آزمایشی به منظور تعیین نوع تابع هسته بهینه صورت گرفته است. تنظیم پارامترهای بهینه این روش و سایر پارامترها در بخش بعد مورد بحث قرار گرفته است.

به منظور بررسی اثربخشی بکارگیری توابع هسته مختلف در KELM، آزمایشی بر روی مجموعه تصویر NUS-WIDE-Object صورت گرفته است. در این آزمایش، روش KELM با تابع هسته مختلف در گام ۲ روش MLENN-GKELM بکار گرفته می‌شود. تابع هسته بکار گرفته

شده شامل هسته خطی<sup>۱</sup>، RBF، چند جمله‌ای<sup>۲</sup> و سیگموید<sup>۳</sup> هستند. پارامترهای هر تابع هسته نیز به ازای RBF، چند جمله‌ای و سیگموید به ترتیب برابر ۵، (۱، ۰/۲) و (۱، ۰/۰) در نظر گرفته شده است. از آنجایی که اثربخشی توابع هسته در روش KELM مورد بررسی قرار گرفته است، معیار تصمیم‌گیری، نتایج شرح‌گذاری بکارگیری روش KELM می‌باشد. شکل ۲-۵ نمودار مقایسه عملکرد شرح‌گذاری بکارگیری توابع هسته مختلف در بکارگیری روش KELM در فرآیند انتخاب نمونه‌های اولیه و شرح‌گذاری را نشان می‌دهد.



شکل ۲-۵: اثربخشی بکارگیری توابع هسته مختلف در روش MLENN-GKELM

همانطور که از نمودار شکل ۲-۵ مشاهده می‌شود، بکارگیری تابع هسته RBF دارای بهترین عملکرد شرح‌گذاری می‌باشد. تابع هسته RBF تابع پیشنهادی [۷۰] در KELM می‌باشد. بنابراین در سایر آزمایش‌های صورت گرفته در این پژوهش از این تابع هسته استفاده شده است. تنظیم پارامترهای بهینه روش MLENN-GKELM ارائه شده: در الگوریتم MLENN-GKELM ارائه شده، تعداد نزدیکترین همسایه‌ها ( $K$ )، پارامترهای الگوریتم پیش‌بینی KELM و  $\theta_1$  در دو حالت بکارگیری روش حریصانه و روش KELM باستی تنظیم شوند.

<sup>1</sup> Linear

<sup>2</sup> Polynomial

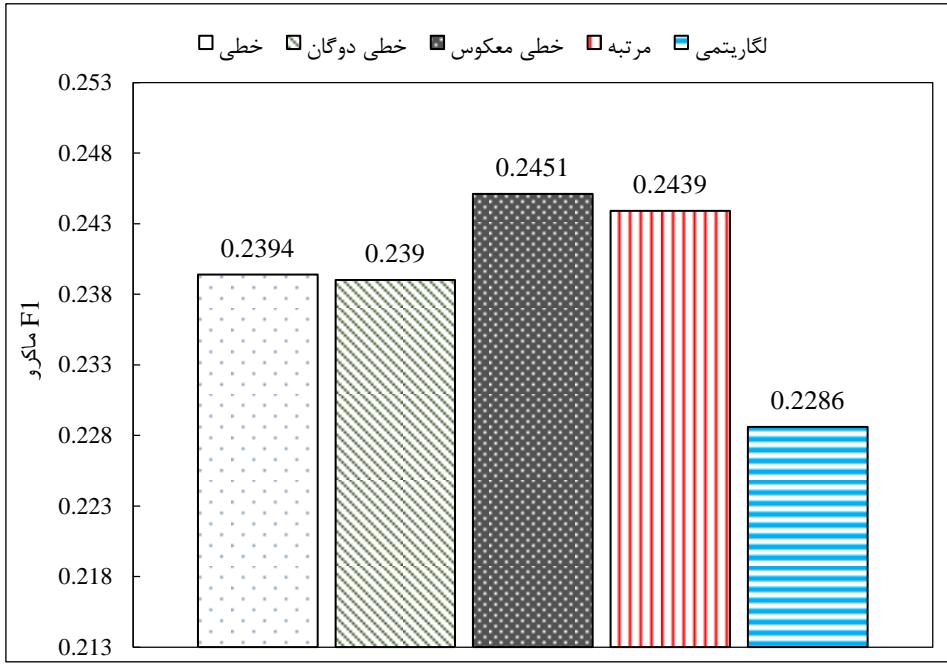
<sup>3</sup> Sigmoid

در هر مرحله از تنظیم، سایر پارامترها را ثابت در نظر گرفته و پارامتر مورد نظر را تغییر داده و مقدار مربوط به بهترین نتیجه به عنوان پارامتر بهینه در نظر گرفته می‌شود. بدین منظور مقادیر  $\theta_1$  در بازه  $0$  تا  $1$ ،  $K$  در بازه  $5$  تا  $525$ ،  $C$  در بازه  $2^{-5}$  تا  $2^0$  و پارامتر تابع هسته RBF در بازه  $1$  تا  $10$  در نظر گرفته شده است. از آنجا که رفتار مجموعه تصاویر چندبرچسبه استفاده شده در برخی از آزمایش‌ها مشابه می‌باشد، آزمایش مجددی صورت نگرفته و نتایج آزمایش‌ها در مجموعه NUS-WIDE-Object به عنوان پارامتر بهینه در این مجموعه تصاویر در نظر گرفته شده است. به عبارت دیگر تغییر در پارامترها تغییری در نتیجه اعمال الگوریتم ایجاد نخواهد کرد. مقادیر بهینه پارامترهای تنظیم شده بر اساس هر مجموعه تصویر به تفکیک در جدول ۱-۵ نشان داده شده است.

## ۲-۵-۵ تنظیم‌های شرح‌گذاری

در این قسمت تنظیم‌های مربوط به شرح‌گذاری مورد بحث و بررسی قرار می‌گیرند. در همه تنظیم‌های صورت گرفته در این بخش در ابتدا مجموعه آموزش کاهش یافته و سپس برای تنظیم پارامتر شرح‌گذاری، به روش اعتبارسنجی متقابل دو بخشی داده شده و پارامتر بهینه یافت می‌شود. در ادامه این قسمت نحوه تنظیم‌های مربوطه به تفکیک بیان شده است.

**بررسی اثربخشی بکارگیری روابط وزن‌دهی متعدد در KNN وزن‌دار:** به منظور بررسی اثربخشی رویکردهای مختلف وزن‌دهی در الگوریتم KNN وزن‌دار در شرح‌گذاری، آزمایشی صورت گرفته است. در این آزمایش اثربخشی  $5$  رویکرد مختلف وزن‌دهی معرفی شده در بخش  $4-4$  مورد بررسی قرار گرفته است. بدین منظور هر رویکرد وزن‌دهی در روش KNN وزن‌دار با استفاده از اعتبارسنجی متقابل دوبخشی روی مجموعه آموزش NUS-WIDE-Object اجرا و ماکزیمم میانگین دقت شرح‌گذاری در این بازه بدست آمده است. شکل ۳-۵ نمودار مقایسه دقت شرح‌گذاری با بکارگیری رویکردهای مختلف وزن‌دهی در روش شرح‌گذاری KNN وزن‌دار را نشان می‌دهد.



شکل ۵-۳: اثربخشی بکارگیری رویکردهای مختلف وزن دهی در روش شرح‌گذاری KNN وزن دار.

همانطور که از نمودار شکل ۵-۳ مشاهده می‌شود، روش وزن دهی خطی معکوس دارای بالاترین مقدار F1 ماکرو است. این رویکرد وزن دهی، رویکرد پایه‌ای ارائه شده در [۶۶] است. با توجه به آزمایش‌های صورت گرفته موثرترین رویکرد وزن دهی نمونه‌ها در الگوریتم KNN وزن دار در کاربرد شرح‌گذاری تصاویر در مقیاس بزرگ نیز همین رویکرد می‌باشد. بنابراین در سایر آزمایش‌های صورت گرفته در این پژوهش از این رویکرد استفاده شده است.

**تنظیم پارامترهای بهینه شرح‌گذاری:** همانطور که بیان شد، در شرح‌گذاری تصاویر بدون برچسب از دو روش KNN وزن دار و KELM استفاده شده است. پارامتر روش KNN وزن دار، تعداد نزدیکترین همسایه‌ها ( $K$ ) و پارامترهای روش KELM، ضریب تاثیر و پارامترتابع هسته RBF است. همچنین دو پارامتر  $\theta_2$  و  $\theta_3$  که نشان دهنده حد آستانه برای تصمیم‌گیری چندبرچسبه هستند بایستی تنظیم شوند. بدین منظور مقدار  $K$  در بازه ۱ تا ۱۰۰، مقدار  $C$  در بازه ۵ تا ۲۰ و پارامترتابع هسته RBF در بازه ۱ تا ۱۰ در نظر گرفته شده است. همچنین مقادیر  $\theta_2$  و  $\theta_3$  نیز در بازه ۰ تا ۱ تنظیم شده است. به مانند آزمایش‌های قبل، تنظیم این پارامترها نیز بر اساس روش اعتبارسنجی متقابل دوبخشی صورت گرفته است. مقادیر بهینه تنظیم شده پارامترهای شرح‌گذاری در جدول ۱-۵ آمده است.

جدول ۱-۵) پارامترهای بهینه تنظیم شده انتخاب نمونه‌های اولیه و شرح‌گذاری.

روش MLENN-GKELM					توضیحات	روش COCNN	روش	انتخاب نمونه‌های اولیه	
Nus	Lite	Scene	Object			Object			
۲۵	۲۵	۲۵	۲۵	$K$	رابطه عدم شباهت دودویی بهینه، و تری	۲	$K$		
۰/۷۵	۰/۷۵	۰/۷۵	۰/۷۵	$\Theta_1$	هسته بهینه در روش RBF، KELM با پارامتر بهینه تنظیم شده ۲/۵	۱۰۰	$\alpha$		
۲۱	۲۱	۲۱	۲۱	$C$	پیش‌بینی مبتنی بر KELM	پیش‌بینی مبتنی بر حریصانه	$\beta$		
۶	۶	۶	۶	$K$					
۰/۷۵	۰/۷۵	۰/۷۵	۰/۷۵	$\Theta_1$	رابطه وزن دهی بهینه در روش شرح‌گذاری KNN وزن دار، وزن دهی خطی معکوس	شرح‌گذاری با روش KNN پایه (K=1)	پیش‌گذاری		
۸۹	۳۹	۵۵	۸۷	$K$	هسته بهینه در روش شرح‌گذاری RBF، KELM با پارامتر بهینه تنظیم شده ۴/۵				
۰/۵	۰/۶۵	۰/۵۵	۰/۳		در روش KNN وزن دار				
۰/۰۷	۰/۰۵	۰/۰۸	۰/۱۸						

## ۶-۵- ارائه، تحلیل و ارزیابی نتایج تجربی

در این قسمت به بیان نتایج تجربی حاصل از اعمال دو رویکرد COCNN و

GKELM ارائه شده و همچنین شرح‌گذاری تصاویر بدون برچسب با الگوریتم‌های شرح‌گذاری

طرح شده در هر مجموعه تصویر می‌پردازیم. در همه آزمایش‌های صورت گرفته برای هر

مجموعه، نتایج کاهش حجم و انتخاب نمونه‌های اولیه و همچنین نتایج طبقه‌بندی در هر مجموعه

مورد مقایسه قرار گرفته است.

در آزمایش‌های صورت گرفته، در ابتدا مجموعه آموزش با رویکرد انتخاب نمونه‌های اولیه

ارائه شده کاهش یافته و سپس مجموعه تصاویر بدون برچسب آزمایش با بکارگیری مجموعه

آموزش کاهش یافته شرح‌گذاری می‌شود. همانطور که بیان شد در بکارگیری روش‌های PS سه

دسته از نتایج دارای اهمیت است. عملکرد شرح‌گذاری با معیارهای اندازه‌گیری دقت، عملکرد

کاهش حجم با معیار نرخ کاهش و تعادل مابین دقت و کاهش حجم با نرخ تعادل ماکرو و میکرو

سه معیار با اهمیت در مقایسه نتایج هستند. در کلیه آزمایش‌های صورت گرفته، این سه دسته

معیار محاسبه و مورد مقایسه قرار گرفته است. همچنین معیار زمان شرح‌گذاری هر تصویر بدون برچسب نیز محاسبه و ارائه شده است. این معیار نشان دهنده مقیاس‌پذیری هر روش می‌باشد.

مقایسه مابین نتایج بدست آمده با دو رویکرد ارائه شده است. در ابتدا، نتایج انتخاب نمونه‌های اولیه ایده‌ای شده با سایر روش‌های پایه‌ای بکار گرفته شده در این پایان‌نامه مورد مقایسه قرار گرفته است. بدین منظور برای هر مجموعه تصویر، نتایج اعمال ایده بکارگیری رابطه عدم شباهت چندبرچسبی مناسب (وتری)، ایده پیش‌بینی حریصانه و رابطه عدم شباهت چند برچسبی مناسب و همچنین ایده پیش‌بینی KELM و رابطه عدم شباهت چند برچسبی مناسب، با یکدیگر و روش MLENN پایه مورد مقایسه قرار گرفته است. به منظور شرح‌گذاری تصاویر بدون برچسب آزمایش نیز از روش پایه شرح‌گذاری (K=1) KNN و دو روش شرح‌گذاری KNN وزن‌دار و KELM بیان شده، استفاده شده است. در رویکرد دوم ارائه نتایج، روش ارائه شده با بکارگیری ایده‌های کاهش حجم و شرح‌گذاری با سایر روش‌های شرح‌گذاری مجموعه تصاویر مقیاس بزرگ ارائه شده مورد مقایسه قرار گرفته است. لازم به ذکر است روش‌های مورد مقایسه در رویکرد دوم هیچ گونه کاهش حجمی بر روی مجموعه آموزش صورت نمی‌دهند و چالش اصلی آنها دقت می‌باشد. از این نظر روش ارائه شده دارای برتری است. در ادامه این بخش نتایج بدست آمده به تفکیک هر مجموعه تصویر آورده شده است.

## ۱-۶-۵ نتایج NUS-WIDE-Object

در این بخش نتایج ارزیابی روش‌های ارائه شده بر روی مجموعه NUS-WIDE-Object ارائه شده است. در این مجموعه تصویر دو رویکرد ارائه شده در انتخاب نمونه‌های اولیه و همچنین روش‌های شرح‌گذاری اجرا و نتایج بدست آمده با یکدیگر مورد مقایسه قرار گرفته است. همانطور که بیان شد، مقایسه در دو حوزه عملکرد انتخاب نمونه‌های اولیه و عملکرد طبقه‌بندی صورت گرفته است. جدول ۲-۵ مقایسه نتایج و عملکرد رویکردهای ارائه شده در سه بخش انتخاب نمونه‌های اولیه و کاهش حجم، شرح‌گذاری و نرخ تعادل مابین کاهش حجم و شرح‌گذاری را در مجموعه NUS-WIDE-Object نشان می‌دهد.

جدول ۲-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Object.

زمان شرح‌گذاری (ثانیه)	نرخ تعادل شرح‌گذاری / کاهش حجم		عملکرد کاهش حجم		عملکرد شرح‌گذاری						روش
	نرخ تعادل	نرخ تعادل	نرخ کاهش	F1	فراخوان	دقت	F1	فراخوان	دقت	دقت	
	میکرو (%)	ماکرو (%)	کاهش (%)	میکرو (%)	میکرو (%)	میکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	
۰/۰۵۶۵	۳/۷۴	۳/۱۱	۱۲/۷۳	۲۹/۳۸	۳۰/۱۹	۳۰/۷۱	۲۴/۴	۲۴/۷	۲۴/۱۱	[۲۲] <sup>۱</sup> CNN	
۰/۰۲۹۹	۱۹/۱۹	۱۴/۹۵	۶۱/۳۳	۳۱/۲۹	۳۰/۰۹	۳۲/۵۸	۲۴/۳۷	۲۳/۲۷	۲۵/۵۹	[۱۲] <sup>۱</sup> MLENN	
۰/۰۲۳	۲۰/۹۹	۱۷/۴۵	۷۰/۶۰	۲۹/۷۳	۳۰/۱۵	۲۹/۳۳	۲۴/۷۲	۲۴/۵۵	۲۴/۸۹	<sup>۱</sup> MLENN(Chord)	
۰/۰۳۸۵	۱۴/۵۵	۱۲/۸۱	۵۱/۸۶	۲۸/۰۶	۲۸/۰۳	۳۰/۴۲	۲۴/۷۱	۲۴/۵۰	۲۴/۹۲	<sup>۱</sup> COCNN	
۰/۰۲۱۴	۲۱/۶۹	۱۸/۰۸	۷۲/۷۳	۲۹/۸۳	۲۸/۸۳	۲۹/۸۲	۲۴/۸۶	۲۴/۳۹	۲۵/۳۴	<sup>۱</sup> MLENN-GKELM(G+KNN)	
۰/۰۲۱۴	۲۲/۷۱	۱۹/۱۹	۷۲/۷۳	۳۱/۲۲	۲۷/۶۷	۳۵/۸۲	۲۶/۳۹	۲۵/۰۲	۲۷/۹۲	<sup>۱</sup> MLENN-GKELM(G+WKNN)	
۰/۰۱۹۸	۲۲/۶۶	۱۸/۸۸	۷۴/۱۵	۳۰/۵۶	۲۹/۷۱	۳۱/۴۵	۲۵/۵۰	۲۵/۲۱	۲۵/۷۰	<sup>۱</sup> MLENN-GKELM(KELM+KNN)	
۰/۰۱۹۸	۳۰/۴۵	۲۶/۹۲	۷۴/۱۵	۴۱/۰۶	۴۱/۰۵	۴۱/۰۷	۲۶/۳۱	۳۸/۹۰	۲۴/۰۵	<sup>۱</sup> MLENN-GKELM(KELM+KELM)	

۱. انتخاب نمونه‌های اولیه + شرح‌گذاری با روش KNN (K=1)  
 ۲. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب + شرح‌گذاری با روش KNN (K=1)  
 ۳. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN (K=1)  
 ۴. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN وزن دار  
 ۵. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KNN (K=1)  
 ۶. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KELM

در جدول ۲-۵ نتایج مربوط به ایده‌های ارائه شده در این پایان‌نامه بصورت پرنگ نشان

داده شده است. در این مقایسه اثربخشی هر ایده بصورت جداگانه مورد بررسی قرار گرفته است. به

عبارت دقیق‌تر، روش COCNN ارائه شده با روش پایه CNN مقایسه شده و اثربخشی ایده حذف

نمونه‌های پرت طبقه‌ای در این روش نشان داده شده است.

در روش MLENN-GKELM ارائه شده، اثربخشی هر ایده بصورت جداگانه با روش

MLENN پایه و MLENN با انتخاب رابطه عدم شباهت دودویی مناسب (وتری) مقایسه شده

است. در نتایج ارائه شده در جدول ۲-۵، MLENN-GKELM(G+KNN) به معنی بکارگیری

رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه در MLENN-GKELM و MLENN-KNN (K=1)

پایه در شرح‌گذاری است. MLENN-GKELM(G+WKNN) به معنی بکارگیری رابطه

عدم شباهت دودویی مناسب و پیش‌بینی حریصانه در MLENN-GKELM و KNN وزن دار در

شرح‌گذاری است. MLENN-GKELM(KELM+KNN) به معنی بکارگیری رابطه عدم شباهت

دودویی مناسب و پیش‌بینی KELM در KELM و KNN پایه در

شرح‌گذاری است. در نهایت MLENN-GKELM(KELM+KNN) به معنی بکارگیری رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM در MLENN-GKELM و KELM در شرح‌گذاری است.

همانطور که در جدول ۲-۵ مشاهده می‌شود، روش CNN مقدار کاهش حجمی برابر ۱۲/۷۳٪ دارد. با بکارگیری ایده حذف نمونه‌های پرت طبقه‌ای و ارائه روش COCNN در این پژوهش، مقدار کاهش حجم مجموعه آموزش به ۰۵۱/۸۶٪ رسیده است. مقدار کاهش حجم بدست آمده نسبت به تغییر دقت به مراتب بیشتر است. به بیان دقیق‌تر، بهبود نرخ تعادل ماکرو در روش COCNN نسبت به روش CNN پایه ۹/۷٪ می‌باشد. این مقدار نشان دهنده موثر بودن ایده بکارگرفته شده در روش CNN پایه است. از آنجا که در آزمایش‌های صورت گرفته به طور کلی زمان یادگیری و نسبت کاهش دقت به کاهش حجم روش CNN زیاد است، این روش در سایر مجموعه‌ها پیاده‌سازی نشده و صرفاً اثربخشی بکارگیری ایده حذف نمونه‌های پرت طبقه‌ای در این روش در مجموعه NUS-WIDE-Object نشان داده شده است.

با توجه به نتایج ارائه شده در جدول ۲-۵، رویکرد MLENN مقدار کاهش حجمی برابر ۳۳/۶۱٪ دارد. با تغییر ایده ارائه شده در [۱۲] با بکارگیری رابطه عدم شباهت دودویی و تری در MLENN(Chord)، بهبود ۲۷/۹٪ در کاهش حجم نسبت به روش MLENN حاصل شده است. (تغییر صورت گرفته در این روش بخشی از ایده اثربخشی بکارگیری روابط عدم شباهت دودویی مختلف بررسی شده در این پژوهش می‌باشد) همچنین با بکارگیری ایده پیش‌بینی برچسب با MLENN و MLENN-GKELM(G+KNN) و استفاده از روش‌های KELM و GKELM(KELM+KNN) روش حریصانه و استفاده از صلاحیت بکارگیری ایده‌های ارائه شده در جهت کاهش حجم مجموعه است. این نتایج نشان از صلاحیت بکارگیری ایده‌های ارائه شده در جهت کاهش حجم مجموعه آموزش و انتخاب نمونه‌های اولیه دارد.

بکارگیری روش KNN (K=1) پایه در شرح‌گذاری برای نشان دادن نحوه انتخاب نمونه‌های موثر در روش ارائه شده نسبت به سایر روش‌ها می‌باشد. این مطلب در عملکرد شرح‌گذاری دو روش MLENN-GKELM(KELM+KNN) و MLENN-GKELM(G+KNN) نسبت به

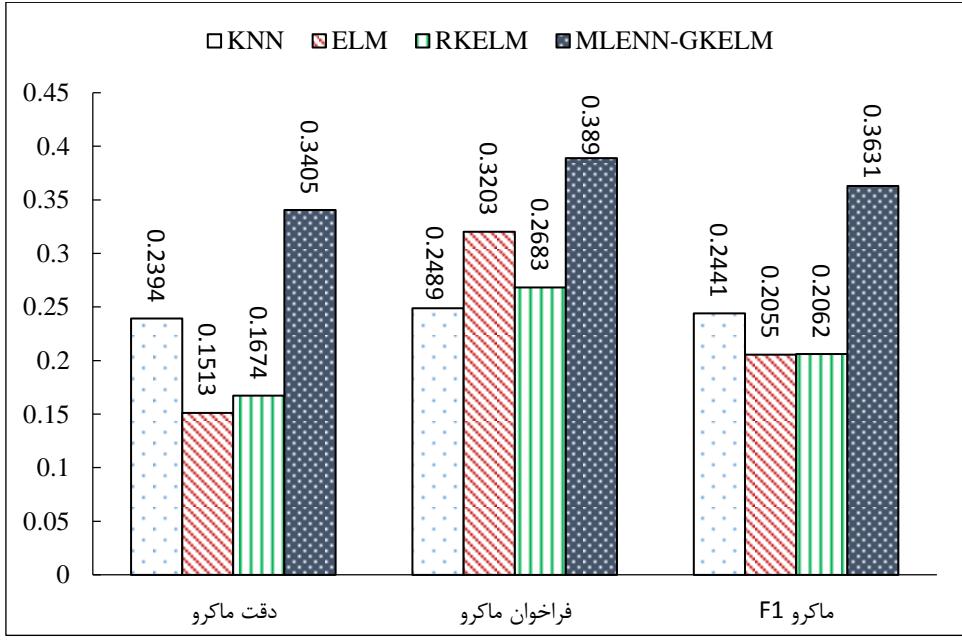
MLENN قابل مشاهده است. همانطور که مشاهده می‌شود، مقدار F1 ماکرو در روش MLENN برابر ۲۴/۳۷٪ است. بکارگیری ایده‌های مطرح شده نه تنها منجر به بهبود نرخ کاهش شده است، بلکه بهبود دقت شرح‌گذاری را نیز به همراه داشته است. همانطور که مشاهده می‌شود، مقدار F1 ماکرو برای این دو روش به ترتیب برابر ۲۴/۸۶٪ و ۲۵/۵٪ است. بهبود دقت بدست آمده نشان دهنده انتخاب نمونه‌های موثرتر روش‌های ارائه شده است.

همانطور که در فصل قبل بیان شد، ایده این پایان‌نامه در جهت بهبود دقت شرح‌گذاری، بکارگیری الگوریتم‌های یادگیری مناسب و با دقت بالا است. دو روش KNN وزن‌دار و KELM به همراه ایده‌های کاهش حجم در دو روش MLENN-GKELM(G+WKNN) و MLENN- GKELM(KELM+KELM) بهبود دقت بالایی نسبت به روش پایه داشته است. بهبود F1 ماکرو در این دو روش نسبت به MLENN به ترتیب برابر ۲/۰۲٪ و ۱/۹۴٪ است.

در نرخ تعادل شرح‌گذاری و کاهش حجم نیز اثربخش بودن ایده‌های ارائه شده قابل مشاهده است. بهبود نرخ تعادل ماکرو در چهار روش MLENN-، MLENN-GKELM(G+KNN)، MLENN-GKELM(KELM)، MLENN-GKELM(KELM+KNN)، GKELM(G+WKNN) + نسبت به روش MLENN به ترتیب برابر ۳/۱۳٪، ۴/۲۴٪، ۳/۹۳٪ و ۱/۹۷٪ می‌باشد.

بهبود نرخ میکرو در این سه روش نیز نسبت به روش MLENN قابل مشاهده است. کاهش حجم مجموعه آموزش زمان شرح‌گذاری را کاهش می‌دهد. کاهش حجم با بکارگیری ایده‌های پیشنهادی، زمان شرح‌گذاری هر تصویر را بهبود داده است. بهبود زمان شرح‌گذاری در چهار روش MLENN-، MLENN-GKELM(G+WKNN)، MLENN-GKELM(G+KNN)، GKELM(KELM+KNN) نسبت به روش MLENN-GKELM(KELM + KELM)، GKELM(KELM+KNN) به ترتیب برابر ۰/۰۰۳۱، ۰/۰۰۱۵، ۰/۰۰۱۵ و ۰/۰۰۳۱ ثانیه می‌باشد. بنابراین با بکارگیری ایده کاهش حجم، مشکل مقیاس‌پذیری و زمان شرح‌گذاری تصاویر در سیستم‌های مقیاس بزرگ مرتفع می‌گردد. همچنین با تعداد کم تصاویر آموزش، مشکل روش‌های یادگیری که با تعداد بالای نمونه‌ها قادر به انجام فرآیند طبقه‌بندی نیستند نیز حل می‌شود.

نتایج شرح‌گذاری تصاویر بدون برچسب آزمایش با بکارگیری روش MLENN-NUS-WIDE با سایر روش‌های شرح‌گذاری در مجموعه تصویر GKELM(KELM + KELM) مورد مقایسه قرار گرفته است. در این مقایسه روش‌های KNN (K=1) پایه [۵]، ELM [۶] و RKELM [۷] پیاده‌سازی و نتایج شرح‌گذاری آن‌ها محاسبه شده است. در تمامی روش‌های مورد مقایسه از تعداد تصاویر آموزش و آزمایش و ویژگی‌های مشابه با این پژوهش استفاده شده است. شکل ۴-۵ نمودار مقایسه‌ای روش MLENN-GKELM(KELM + KELM) با سایر روش‌ها را در مجموعه تصویر NUS-WIDE-Object نشان می‌دهد.



شکل ۴-۵: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE-Object

همانطور که نمودار شکل ۴-۵ نشان می‌دهد، عملکرد شرح‌گذاری روش MLENN-GKELM ارائه شده از سایر روش‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ بهتر می‌باشد. بهبود عملکرد این روش نسبت به روش‌های KNN (K=1) پایه، ELM و RKELM در معیار F1 ماکرو به ترتیب برابر ۱۱٪/۹، ۱۵٪/۶۹ و ۱۵٪/۷۶ می‌باشد. این در حالی است که بر اساس نتایج ارائه شده در جدول ۲-۵ علاوه بر دقت بالای شرح‌گذاری، روش ارائه شده عملکرد خوبی در کاهش حجم مجموعه آموزش نیز دارا می‌باشد. از این‌رو این روش صلاحیت خوبی در بکارگیری در محیط‌های مقیاس بزرگ و واقعی دارد.

## ۲-۶-۵ نتایج NUS-WIDE-Scene

در این بخش نتایج ارزیابی روش ارائه شده بر روی مجموعه NUS-WIDE-Scene مورد بررسی قرار می‌گیرد. نتایج این بخش نیز به مانند نتایج ارائه شده در بخش قبل می‌باشد. جدول ۳-۵ مقایسه نتایج و عملکرد رويکردهای ارائه شده در سه بخش انتخاب نمونه‌های اولیه و کاهش NUS-WIDE-Scene، شرح‌گذاری و نرخ تعادل مابین کاهش حجم و شرح‌گذاری را در مجموعه NUS-WIDE-Scene نشان می‌دهد.

همانطور که در جدول ۳-۵ مشاهده می‌شود، رویکرد MLENN کاهش حجمی معادل ۲۱/۳۴٪ را در مجموعه NUS-WIDE-Scene دارد. در صورتی که در (MLENN(Chord)، بهبود ۴۰/۳۶٪ در کاهش حجم نسبت به روش MLENN حاصل شده است. از طرفی با بکارگیری ایده MLENN-پیش‌بینی برچسب با روش حریصانه و KELM و استفاده از روش‌های GKELM(G+KNN) بهبود ۴۶/۳۹٪ و ۳۹/۳۴٪ نسبت به MLENN حاصل شده است. این نرخ کاهش حجم بالا در این روش‌ها نشان از صلاحیت ایده‌های ارائه شده در انتخاب نمونه‌های اولیه دارد.

**جدول ۳-۵** مقایسه عملکرد رويکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Scene

زمان شرح‌گذاری (ثانیه)	عملکرد شرح‌گذاری										روش	
	عملکرد کاهش حجم					عملکرد شرح‌گذاری						
	نرخ تعادل شرح‌گذاری / کاهش حجم	عملکرد کاهش حجم	نرخ تعادل	نرخ تعادل	نرخ	F1	فراخوان	دقت	F1	فراخوان	دقت	
M	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MLENN
٪	٪	٪	٪	٪	٪	٪	٪	٪	٪	٪	٪	MLENN(Chord)
۰/۰۵۷۱	۱۰/۶۵	۵/۲۴	۲۱/۳۴	۴۹/۸۹	۴۶/۶۱	۵۳/۶۶	۲۴/۵۶	۲۲/۵۹	۲۶/۹۱	۱۲/۱	۱۲/۱	
۰/۰۲۸۵	۲۹/۹۲	۱۵/۰۵	۶۱/۷۰	۴۸/۴۹	۴۵/۸۹	۵۱/۴۱	۲۴/۳۹	۲۱/۹۴	۲۷/۴۵	۱۳/۱	۱۳/۱	
۰/۰۲۲۹	۳۲/۳۹	۱۶/۵۱	۶۷/۷۳	۴۹/۳	۴۴/۷۶	۵۴/۸۶	۲۴/۳۷	۲۱/۱۸	۲۸/۶۹	۱۴/۱	۱۴/۱	
۰/۰۲۵	۳۶/۳۹	۱۹/۳۱	۶۷/۷۳	۵۲/۷۳	۴۸/۳	۶۰/۵۵	۲۸/۵۱	۳۰/۱۸	۲۷/۰۱	۱۵/۰	۱۵/۰	
۰/۰۳۱۲	۳۰/۵۷	۱۵/۰۹	۶۰/۶۸	۵۰/۳۷	۴۶/۱۸	۵۵/۴۱	۲۴/۸۶	۲۱/۹۵	۲۸/۶۶	۱۶/۰	۱۶/۰	
۰/۰۳۱۱۸	۳۶/۵	۲۰/۱۷	۶۰/۶۸	۶۰/۱۵	۵۵/۱۱	۶۶/۲۱	۳۲/۲۴	۳۲/۹۰	۳۲/۵۹	۱۷/۰	۱۷/۰	

۱. انتخاب نمونه‌های اولیه + شرح‌گذاری با روش KNN (K=1)

۲. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب + شرح‌گذاری با روش KNN (K=1)

۳. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN (K=1)

۴. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN وزن دار

۵. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KNN (K=1)

۶. انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KELM

از طرفی همانطور که مشاهده می‌شود، بکارگیری ایده‌های مطرح شده نه تنها منجر به بهبود نرخ کاهش شده است، بلکه بهبود دقت شرح‌گذاری را به همراه داشته است. بر اساس جدول ۳-۵، بهبود F1 ماکرو در دو روش MLENN-GKELM(G+WKNN) و MLENN-

است. بهبود نرخ تعادل میکرو در این چهار روش متفاوت است. مثلاً MLENN نسبت به روش GKELM(KELM+KELM) به ترتیب برابر ۰/۳/۹۵ و ۰/۸/۶۸ است.

در نرخ تعادل شرح‌گذاری و کاهش حجم نیز اثربخش بودن ایده‌های ارائه شده قابل مشاهده است. بهبود نرخ تعادل میکرو در چهار روش MLENN-GKELM(G+KNN) و MLENN-

است. بهبود نرخ تعادل میکرو در این چهار روش متفاوت است. مثلاً MLENN نسبت به روش GKELM(KELM + KELM) به ترتیب برابر ۰/۱۴/۰۷ و ۰/۱۴/۰۶٪ است. بهبود نرخ تعادل میکرو در این چهار روش نیز نسبت به روش MLENN به ترتیب برابر ۰/۹/۸۵ و ۰/۱۴/۹۳٪ می‌باشد.

بهبود زمان شرح‌گذاری در چهار روش MLENN به ترتیب برابر ۰/۲۵/۷۴٪، ۰/۲۲/۷۴٪، ۰/۲۵/۷۴٪ و ۰/۲۵/۸۵٪ می‌باشد.

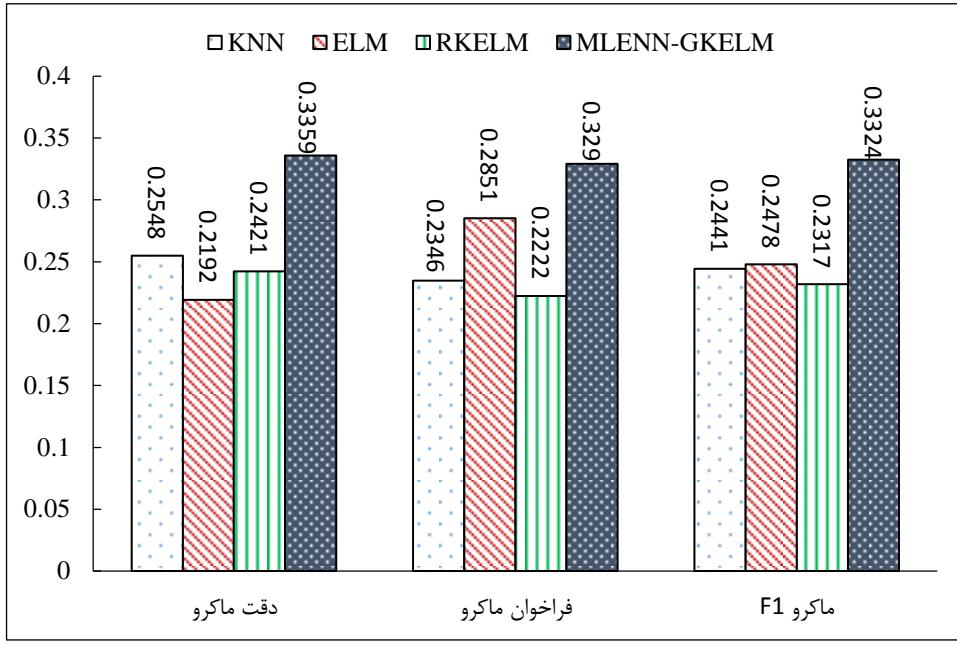
زمان تقریبی شرح‌گذاری هر تصویر بدون برچسب آزمایش در روش MLENN ۰/۰۵/۷۱ ثانیه است. بهبود زمان شرح‌گذاری در چهار روش MLENN-GKELM(G+KNN) و MLENN-

است. بهبود زمان شرح‌گذاری در چهار روش MLENN-GKELM(KELM+KNN) و GKELM(G+WKNN)

به ترتیب برابر ۰/۰۳/۲۱ و ۰/۰۳/۳۲٪ است. بهبود زمان شرح‌گذاری حاصل شده، مقیاس‌پذیر ۰/۰۲۵۹ و ۰/۰۲۵۹٪ می‌باشد.

بر اساس بهبود زمان شرح‌گذاری حاصل شده، مقیاس‌پذیر بودن رویکرد پیشنهادی در این مجموعه تصویر قابل استنتاج است.

به مانند نتایج NUS-WIDE-Object، نتایج شرح‌گذاری تصاویر بدون برچسب آزمایش با سایر روش‌های شرح‌گذاری مورد مقایسه قرار گرفته است. شکل ۴-۵ نمودار مقایسه‌ای روش NUS-WIDE با سایر روش‌ها را در مجموعه تصویر MLENN-GKELM(KELM + KELM) نشان می‌دهد.



شکل ۵-۵: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE-Scene

همانطور که نمودار شکل ۵-۵ نشان می‌دهد، عملکرد شرح‌گذاری روش MLENN-GKELM ارائه شده از سایر روش‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ در مجموعه NUS-WIDE-Scene بهتر است. بهبود عملکرد این روش نسبت به روش‌های KNN (پایه)، ELM و RKELM در معیار F1 ماکرو به ترتیب برابر ۰/۷۰٪، ۰/۴۶٪ و ۰/۱۰٪ می‌باشد. این در حالی است که به مانند مجموعه NUS-WIDE-Object علاوه بر دقت بالای شرح‌گذاری، روش ارائه شده عملکرد خوبی در کاهش حجم مجموعه آموزش نیز دارا می‌باشد. از این‌رو این روش صلاحیت خوبی در حل چالش مقیاس‌پذیری و دقت در محیط‌های مقیاس بزرگ و واقعی دارد.

### ۳-۶-۵ نتایج NUS-WIDE-Lite

در این بخش نتایج ارزیابی بر روی مجموعه NUS-WIDE-Lite ارائه شده است. نتایج این بخش نیز به مانند نتایج ارائه شده در دو بخش قبل می‌باشد. جدول ۴-۵ مقایسه نتایج و عملکرد رویکردهای ارائه شده در سه بخش انتخاب نمونه‌های اولیه و کاهش حجم، شرح‌گذاری و نرخ تعادل مابین کاهش حجم و شرح‌گذاری را در مجموعه NUS-WIDE-Lite نشان می‌دهد.

همانطور که جدول ۴-۵ نشان می‌دهد، در رویکرد MLENN(Chord) بهبود ۴۹/۵٪ در کاهش حجم نسبت به روش MLENN حاصل شده است. از طرفی با بکارگیری ایده پیش‌بینی MLENN-GKELM(G+KNN) در روشهای KELM و MLENN-GKELM(G+KNN) به ترتیب بهبود ۴۳/۴۳٪ و ۰۵۴/۷۹٪ در کاهش حجم حاصل شده است. این نرخ کاهش حجم بالا در این روش‌ها نشان از برتری رویکردهای ارائه شده در انتخاب نمونه‌های اولیه دارد.

**جدول ۴-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE-Lite**

زمان شرح‌گذاری (ثانیه)	نرخ تعادل شرح‌گذاری / کاهش حجم		عملکرد کاهش حجم		عملکرد شرح‌گذاری						روش
	نرخ تعادل میکرو (٪)	نرخ تعادل ماکرو (٪)	نرخ کاهش (٪)	F1 میکرو (٪)	فراخوان میکرو (٪)	دقت میکرو (٪)	F1 ماکرو (٪)	فراخوان ماکرو (٪)	دقت ماکرو (٪)		
	نرخ تعادل میکرو (٪)	نرخ تعادل ماکرو (٪)	نرخ کاهش (٪)	F1 میکرو (٪)	فراخوان میکرو (٪)	دقت میکرو (٪)	F1 ماکرو (٪)	فراخوان ماکرو (٪)	دقت ماکرو (٪)		
۰/۰۸۷۶	۱۱/۴۲	۴/۵۸	۲۴/۳۰	۴۷	۴۴/۴۱	۴۹/۹۲	۱۸/۸۴	۱۸/۱۵	۱۹/۶۰	[۱۲] <sup>۱</sup> MLENN	
۰/۰۳۱۶	۳۴/۶۰	۱۲/۲۲	۷۳/۸۰	۴۶/۸۸	۴۷/۲۳	۴۶/۵۳	۱۷/۹۱	۱۷/۹۴	۱۷/۸۹	<sup>۱</sup> MLENN(Chord)	
۰/۰۲۱۷	۳۸/۹۶	۱۴/۴۲	۶۷/۷۳	۴۷/۸۷	۴۴/۴۱	۵۱/۹۲	۱۷/۷۳	۱۶/۱۲	۱۹/۶۹	<sup>۱</sup> MLENN-GKELM(G+KNN)	
۰/۰۲۲۱	۴۱/۰۸	۱۷/۲۷	۶۷/۷۳	۵۰/۴۸	۴۲/۱۹	۶۲/۸۲	۲۱/۲۲	۲۰/۲۹	۲۲/۲۳	<sup>۱</sup> MLENN-GKELM(G+WKNN)	
۰/۰۱۹۲	۳۸/۲۱	۱۴/۹۹	۷۹/۰۹	۴۸/۳۱	۴۳/۵۲	۵۴/۲۷	۱۸/۹۶	۱۷/۱۶	۲۱/۱۷	<sup>۱</sup> MLENN-GKELM(KELM+KNN)	
۰/۰۱۹۶	۴۴/۲۷	۲۰/۲۷	۷۹/۰۹	۵۵/۸۰	۴۶/۹۶	۶۹/۲۹	۲۵/۶۳	۲۳/۳۶	۲۸/۳۷	<sup>۱</sup> MLENN-GKELM(KELM+KELM)	

۱. انتخاب نمونه‌های اولیه + شرح‌گذاری با روش KNN (K=1)
۲. انتخاب نمونه‌های اولیه با ایده رابطه عدم‌شناخت دودویی مناسب + شرح‌گذاری با روش KNN (K=1)
۳. انتخاب نمونه‌های اولیه با ایده رابطه عدم‌شناخت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN (K=1)
۴. انتخاب نمونه‌های اولیه با ایده رابطه عدم‌شناخت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش KNN وزن دار (K=1)
۵. انتخاب نمونه‌های اولیه با ایده رابطه عدم‌شناخت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KNN (K=1)
۶. انتخاب نمونه‌های اولیه با ایده رابطه عدم‌شناخت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش KELM

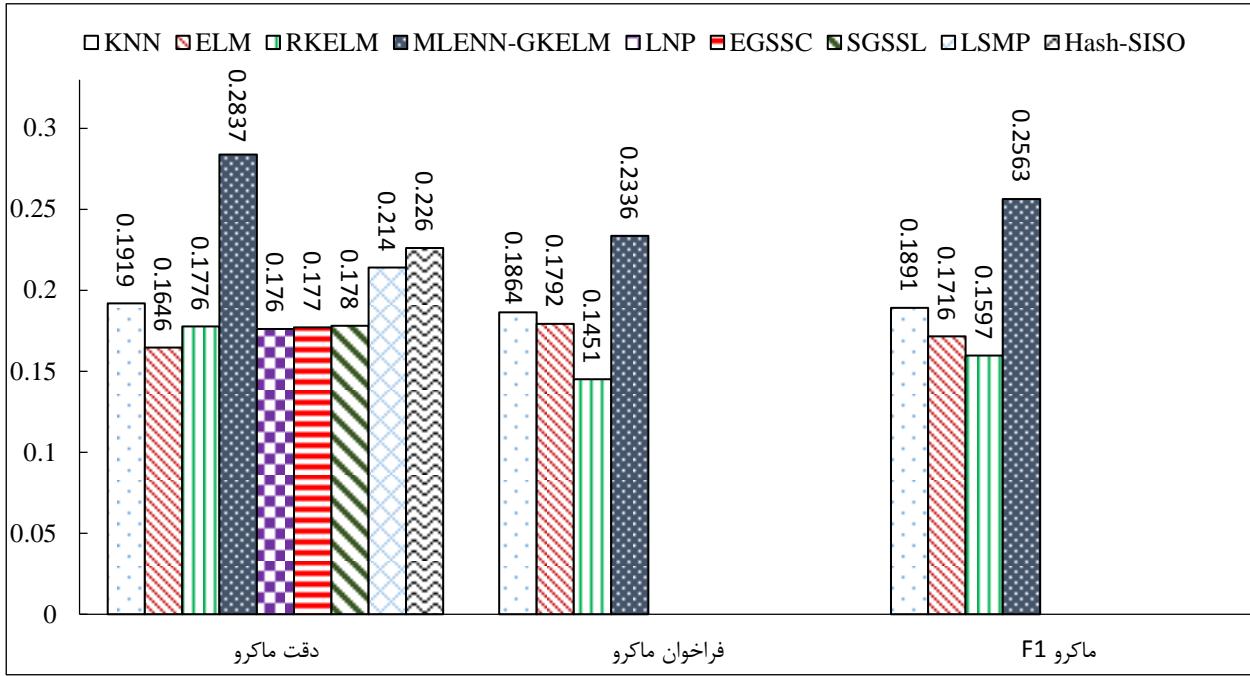
از طرفی همانطور که جدول ۴-۵ نشان می‌دهد، بهبود F1 ماکرو در دو روش MLENN- MLENN-GKELM(KELM+KELM) و GKELM(G+WKNN) نسبت به روش پایه MLENN به ترتیب برابر ۰/۳۸٪ و ۰/۷۹٪ است.

در نرخ تعادل شرح‌گذاری و کاهش حجم نیز اثربخش بودن ایده‌های ارائه شده قابل مشاهده است. بهبود نرخ تعادل ماکرو در چهار روش MLENN- MLENN-GKELM(G+KNN) و MLENN- MLENN-GKELM(KELM+KNN) و GKELM(G+WKNN)

GKELM(KELM + KELM) نسبت به روش MLENN به ترتیب برابر ۹/۸۵٪، ۱۲/۶۹٪، ۱۵/۶۹٪ و ۱۷/۴۱٪ می‌باشد. بهبود نرخ میکرو در این سه روش نیز نسبت به روش MLENN به ترتیب برابر ۲۷/۵۴٪، ۲۹/۶۶٪، ۲۶/۷۹٪ و ۳۲/۸۵٪ می‌باشد.

بهبود زمان شرح‌گذاری در چهار روش MLENN-، MLENN-GKELM(G+KNN) و MLENN-GKELM(KELM+KNN) .GKELM(G+WKNN) می‌باشد. نسبت به روش MLENN به ترتیب برابر ۰/۰۶۵۹، ۰/۰۶۵۵ و ۰/۰۶۸۴ ثانیه می‌باشد.

به مانند نتایج دوبخش قبل، نتایج شرح‌گذاری تصاویر بدون برچسب آزمایش با سایر روش‌های شرح‌گذاری مورد مقایسه قرار گرفته است. در دو مجموعه تصویر NUS-WIDE-Lite و NUS-WIDE-Object نسبت به دو مجموعه NUS-WIDE و NUS-WIDE-Scene صورت گرفته است. از این‌رو در مقایسه‌های صورت گرفته، این پژوهش‌ها نیز گنجانده شده است. نتایج رویکرد ارائه شده با روش‌های (K=1) KNN [۵]، ELM [۴]، HASH- [۲۲] LSMP [۷۳] SGSSL [۷۲] EGSSC [۷۱] LNP [۶] و RKELM [۷۴] SISO مقایسه شده است. از آنجایی که در برخی از مقالات مورد نظر فقط دقت ماکرو در دسترس بوده است، نتایج این معیار با رویکرد ارائه شده مقایسه شده است. شکل ۶-۵ نمودار مقایسه‌ای روش MLENN-GKELM(KELM + KELM) با سایر روش‌ها را در مجموعه تصویر NUS-WIDE-Lite نشان می‌دهد.



شکل ۶-۵: مقایسه عملکرد شرح‌گذاری روشنای مختلف در مجموعه تصویر NUS-WIDE-Lite

همانطور که نمودار شکل ۶-۵ نشان می‌دهد، عملکرد شرح‌گذاری روشنای MLENN-GKELM ارائه شده از سایر روشنای شرح‌گذاری خودکار مجموعه تصاویر در مقایس بزرگ بهتر می‌باشد. این در حالی است که علاوه بر دقت بالای شرح‌گذاری، روشنای ارائه شده عملکرد خوبی در کاهش حجم مجموعه آموزش نیز دارد. از این‌رو این روشن صلاحیت خوبی در بکارگیری در محیط‌های مقایس بزرگ و واقعی دارد. بهبود دقت ماکرو در این روشن نسبت به روشنای KNN (پایه، ELM، RKELM، LSMP، SGSSL، EGSSC، LNP) به ترتیب ۹۱٪، ۷۷٪، ۶۷٪، ۵۷٪، ۴۱٪، ۹۱٪، ۷۷٪، ۶۷٪ و ۵۷٪ می‌باشد.

#### ۶-۶ نتایج NUS-WIDE

در نهایت این بخش به نتایج ارزیابی روشنای ارائه شده بر روی مجموعه NUS-WIDE اختصاص دارد. نتایج این بخش نیز به مانند نتایج ارائه شده در بخش‌های قبلی است. جدول ۵-۵ مقایسه نتایج و عملکرد رویکردهای ارائه شده در سه بخش انتخاب نمونه‌های اولیه و کاهش حجم، شرح‌گذاری و نرخ تعادل مابین کاهش حجم و شرح‌گذاری را در مجموعه NUS-WIDE نشان می‌دهد.

جدول ۵-۵) مقایسه عملکرد رویکردهای ارائه شده در مجموعه تصویر NUS-WIDE

زمان شرح‌گذاری (ثانیه)	نرخ تعادل شرح‌گذاری / کاهش حجم		عملکرد کاهش حجم		عملکرد شرح‌گذاری						روش
	نرخ تعادل	نرخ کاهش	F1	فراخوان	دقت	F1	فراخوان	دقت	ماکرو (%)	ماکرو (%)	
	میکرو (%)	ماکرو (%)	میکرو (%)	میکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	ماکرو (%)	
۰/۲۶۰۵	۳۰/۷۲	۸/۷۸	۵۵/۷۲	۵۵/۱۳	۵۰/۹۴	۶۰/۰۶	۱۵/۷۶	۱۴/۲۱	۱۷/۶۸	۱۲/۱۲	MLENN
۰/۱۹۶۵	۲۱/۷۱	۹/۷۰	۶۷/۲۶	۳۲/۲۸	۳۴/۹۰	۳۰/۰۲	۱۴/۴۳	۱۴/۶۷	۱۴/۱۹	۱۴/۱۹	MLENN(Chord)
۰/۰۶۶۴	۲۸/۱۵	۱۲/۱۳	۸۹/۶۱	۳۱/۴۱	۲۵/۹۷	۳۹/۷۶	۱۳/۵۴	۱۰/۵۷	۱۸/۸۵	۱۸/۸۵	MLENN-GKELM(G+KNN)
۰/۰۶۶۸	۳۷/۲	۱۴/۹۶	۸۹/۶۱	۴۱/۳۱	۳۲/۹۱	۵۲/۸۳	۱۶/۶۹	۱۷/۰۲	۱۶/۳۸	۱۶/۳۸	MLENN-GKELM(G+WKNN)
۰/۰۶۶۰	۲۸/۷۴	۱۳/۱۱	۹۰/۴۳	۳۱/۷۸	۲۴/۷۲	۴۴/۴۹	۱۴/۵۰	۱۰/۷۴	۲۲/۲۸	۲۲/۲۸	MLENN-GKELM(KELM+KNN)
۰/۰۶۶۲	۳۶/۹۲	۲۰/۵۵	۹۰/۴۳	۴۰/۸۴	۳۰/۹۶	۶۰	۲۲/۷۲	۱۹/۵۹	۲۷/۰۵	۲۷/۰۵	MLENN-GKELM(KELM+KELM)

انجام نمونه‌های اولیه + شرح‌گذاری با روش (K=1) KNN (K=1)  
 انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب + شرح‌گذاری با روش (K=1) KNN (K=1)  
 انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش (K=1) KNN (K=1)  
 انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی حریصانه + شرح‌گذاری با روش وزن دار KNN (K=1)  
 انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش (K=1) KELM + شرح‌گذاری با روش (K=1) KELM  
 انتخاب نمونه‌های اولیه با ایده رابطه عدم شباهت دودویی مناسب و پیش‌بینی KELM + شرح‌گذاری با روش (K=1) KELM

همانطور که جدول ۵-۵ نشان می‌دهد، رویکرد MLENN کاهش حجمی برابر ۵۵/۷۲٪

دارد. در صورتی که این کاهش حجم در MLENN(Chord) برابر ۶۷/۲۶٪ است. به عبارت

دقیق‌تر، بهبود کاهش حجم صورت گرفته در MLENN(Chord) نسبت به MLENN برابر

۱۱/۵۴٪ است. با بکارگیری ایده پیش‌بینی برچسب با روش حریصانه و KELM در روش‌های

MLENN-GKELM(KELM+KNN) و MLENN-GKELM(G+KNN) به ترتیب بهبود

۳۳/۸۹٪ و ۳۴/۷۱٪ در کاهش حجم حاصل شده است. این نرخ کاهش حجم بالا در این روش‌ها

نشان از مقیاس‌پذیری رویکردهای ارائه شده دارد.

بهبود F1 ماکرو در دو روش MLENN- و MLENN-GKELM(G+WKNN)

نیز به روش پایه MLENN نسبت به روشنایی GKELM(KELM+KELM) ۰/۰/۹۳٪ و ۰/۰/۹۶٪ است.

است.

در نرخ تعادل شرح‌گذاری و کاهش حجم اثربخش بودن ایده‌های ارائه شده قابل مشاهده

است. بهبود نرخ تعادل ماکرو در چهار روش MLENN- MLENN-GKELM(G+KNN)

MLENN- و MLENN-GKELM(KELM+KNN) و GKELM(G+WKNN)

٪ ۰.۴/۳۳ به ترتیب برابر MLENN نسبت به روش GKELM(KELM + KELM)

و ٪ ۱۱/۷۷ می‌باشد. بهبود نرخ تعادل میکرو نیز در روش کلی ارائه شده قابل مشاهده است.

MLENN- شرح‌گذاری در چهار روش MLENN-GKELM(G+KNN) و GKELM(G+WKNN)

MLENN- ۰/۱۹۴۱ به ترتیب برابر MLENN نسبت به روش GKELM(KELM + KELM)

و ٪ ۰/۱۹۴۵ ثانیه می‌باشد.

به مانند نتایج بخش‌های قبل، نتایج شرح‌گذاری تصاویر بدون برچسب آزمایش با سایر روش‌های شرح‌گذاری مورد مقایسه قرار گرفته است. شکل ۷-۵ نمودار مقایسه‌ای روش-

MLENN-GKELM با سایر روش‌ها در مجموعه تصویر NUS-WIDE نشان می‌دهد.

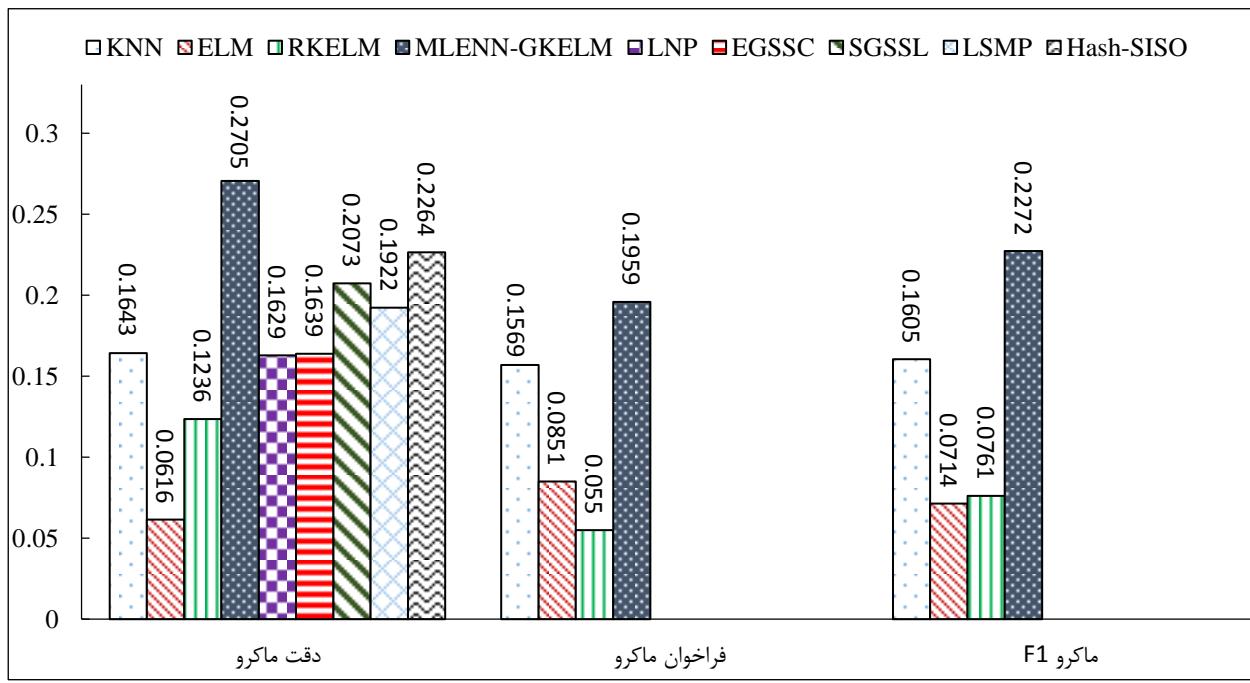
همانطور که در نمودار شکل ۷-۵ مشاهده می‌شود، عملکرد شرح‌گذاری روش-

GKELM ارائه شده از سایر روش‌های ارائه شده در شرح‌گذاری خودکار مجموعه تصاویر در

مقیاس بزرگ بهتر می‌باشد. بهبود عملکرد دقت ماکرو در این روش نسبت به روش‌های KNN

\_HASH-SISO، LSMP، SGSSL، EGSSC، LNP، RKELM، ELM (K=1) پایه، و Hash-SISO به ترتیب

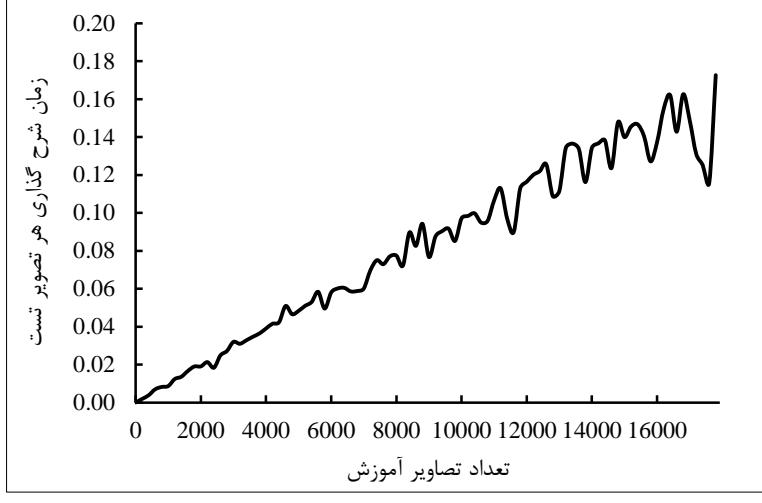
برابر ٪ ۰/۱۰/۶۲، ٪ ۰/۲۰/۸۹، ٪ ۰/۱۴/۶۹، ٪ ۰/۱۰/۷۶، ٪ ۰/۶/۳۲، ٪ ۰/۱۰/۶۶ و ٪ ۰/۴/۴۱ است.



شکل ۷-۵: مقایسه عملکرد شرح‌گذاری روش‌های مختلف در مجموعه تصویر NUS-WIDE

## ۷-۵- تجزیه و تحلیل مقیاس‌پذیری

در این بخش تجزیه و تحلیل مقیاس‌پذیری رویکرد MLENN-GKELM مورد بحث و بررسی قرار می‌گیرد. تجزیه و تحلیل صورت گرفته بر اساس زمان شرح‌گذاری بدست آمده برای هر تصویر بدون برچسب آزمایش است. همانطور که در بخش قبل بیان شد، زمان شرح‌گذاری بدست آمده بر اساس تعداد تصاویر درون مجموعه آموزش متغیر است. در آزمایش صورت گرفته در این بخش تعداد تصاویر درون مجموعه آموزش از ۱ تا ۱۷۹۲۸ تصویر تغییر یافته و زمان شرح‌گذاری هر تصویر بدون برچسب آموزش بدست آمده است. (تغییر تعداد تصاویر بر اساس تعداد تصاویر مجموعه آموزش NUS-WIDE-Object صورت گرفته است). شکل ۸-۵ نمودار زمان شرح‌گذاری را با تغییر تعداد تصاویر درون مجموعه آموزش نشان می‌دهد.



شکل ۸-۵: نمودار زمان شرح‌گذاری هر تصویر بدون برچسب با تغییر تعداد تصاویر مجموعه آموزش.

نمودار شکل ۸-۵ نشان می‌دهد که زمان شرح‌گذاری هر تصویر بدون برچسب آزمایش به تعداد تصاویر مجموعه آموزش وابسته بوده و این زمان بصورت خطی افزایش می‌یابد. به عبارت دقیق‌تر، با مجموعه تصویر آموزش مقیاس بزرگ با حجم بالا زمان زیادی برای شرح‌گذاری هر تصویر صرف می‌شود. در نتیجه پیچیدگی سیستم شرح‌گذاری افزایش می‌یابد. از این‌رو کاهش تعداد تصاویر مجموعه آموزش با بکارگیری ایده‌های مطرح شده در این پایان‌نامه، چالش زمان شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ را حل خواهد کرد.

## جمع‌بندی

در این فصل به بیان نتایج حاصل از پیاده‌سازی رویکردهای پیشنهادی پرداخته شد. این پیاده‌سازی‌ها شامل بررسی اثربخشی بکارگیری روابط عدم شباهت دودویی در گام ۳ از رویکرد پیشنهادی MLENN-GKELM و همچنین بررسی اثربخشی بکارگیری هسته‌های مختلف در روش KELM و بکارگیری مکانیزم‌های متعدد وزن‌دهی در روش KNN وزن‌دار هستند. کلیه آزمایش‌های صورت گرفته در این فصل بر روی مجموعه تصویر NUS-WIDE و نسخه‌های آن شامل Scene و Object Lite انجام شد. رویکردهای ارائه شده در سه بخش عملکرد کاهش حجم، شرح‌گذاری و تعادل مابین کاهش حجم و شرح‌گذاری ارزیابی شد. همچنین به منظور نشان دادن مقیاس‌پذیری رویکردهای ارائه شده، زمان شرح‌گذاری هر تصویر بدون برچسب نیز محاسبه و ارائه شد. نتایج ارائه شده نشان از صلاحیت عملکرد رویکردهای ارائه شده در حل چالش‌های پیش‌روی سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ دارند.

## فصل ۶:

# نتیجه‌گیری و پژوهش‌های آینده

## ۱-۶- مروری بر رویکردهای پیشنهادی

شرحگذاری خودکار تصاویر به معنی توصیف محتوای دیداری تصاویر با استفاده از کلمه‌ها و توصیفگرهای متنی است. به منظور توصیف محتوای دیداری تصاویر، مدل یادگیری از مجموعه تصاویر برچسبدار و بدون برچسب آموزش داده می‌شود. سپس برای هر تصویر بدون برچسب آزمایش، از مدل یادگیری استفاده شده و برچسب‌های تصویر پیش‌بینی می‌شود. با افزایش تعداد تصاویر و همچنین تعداد کلمه‌های شرحگذاری، اندازه مجموعه تصاویر نیز رشد می‌کند. با رشد مجموعه تصاویر، مدل‌های شرحگذاری با چالش‌هایی مواجه می‌شوند. عمدۀ این چالش‌ها مقیاس‌پذیری سیستم یا به عبارتی تضمین بلادرنگ بودن سیستم، زمان شرحگذاری و دقت شرحگذاری هستند.

در فصل دوم از این پایان‌نامه، پژوهش‌های پیشین که چالش‌های پیش‌روی سیستم‌های شرحگذاری خودکار مجموعه تصاویر در مقیاس بزرگ را مورد بررسی قرار داده‌اند، مورد بحث و بررسی قرار گرفت. جدول ۱-۲ خلاصه پژوهش‌های بررسی شده را نشان می‌دهد. روش‌های بررسی شده، چالش‌های پیش‌روی سیستم‌های شرحگذاری خودکار مجموعه تصاویر در مقیاس بزرگ را با سه رویکرد حل کرده‌اند. بکارگیری روش‌های پیش‌پردازش نمونه‌ها، ارائه مدل‌های یادگیری مقیاس‌پذیر و توزیع پذیری محاسبات این سه رویکرد را شامل می‌شوند.

یکی از رویکردهایی که در مقیاس‌پذیر کردن سیستم‌های شرحگذاری با مجموعه تصاویر مقیاس بزرگ می‌توان در نظر گرفت، کاهش حجم نمونه‌های آموزش با بکارگیری روش‌های PS است. PS، فرآیند یافتن نماینده‌های همسایه‌های یک نمونه است. پس از شناسایی و انتخاب نمونه‌های طبقه‌بندی بر اساس نزدیکترین همسایه‌های یک نمونه این اتفاق می‌افتد. بکارگیری این ایده منجر به کاهش حجم مجموعه آموزش و با حجمی کمتر به دست می‌آید. بکارگیری این ایده طبقه‌بندی می‌شود. نتیجه این کار کاهش زمان طبقه‌بندی و افزایش دقت طبقه‌بندی است. همچنین با بکارگیری روش‌های PS در سیستم‌های مقیاس بزرگ و دارای داده‌های آموزش بسیار

زیاد، مشکل روش‌های مبتنی بر نمونه‌ها، که نیازمند حجم بالای ذخیره‌سازی هستند مرتفع می‌شود. در فصل سوم، روش‌های PS مورد بررسی قرار گرفت و طبقه‌بندی از ویژگی‌های آن‌ها بررسی شد. طبقه‌بندی صورت گرفته در شکل ۱-۳ نشان داده شده است.

در فصل چهارم، دو رویکرد انتخاب نمونه‌های اولیه مبتنی بر روش‌های CNN و ENN مطرح شد. در رویکرد اول ارائه شده، روش پایه CNN با ایده شناسایی و حذف نمونه‌های پرت طبقه‌ای تغییر یافت. روش COCNN مطرح شده در نرخ کاهش نمونه‌ها نسبت به روش CNN بهبود داده شد. نتایج بهبود صورت گرفته در جدول ۲-۵ قابل مشاهده است. همانطور که بیان شد، روش CNN از سرعت پایین و دقت کمتری نسبت به روش ENN برخوردار است. از این‌رو ایده‌های مطرح شده بعدی بر روی روش ENN پیاده‌سازی شد.

فرض اصلی روش‌های پایه PS از جمله ENN به تک برچسب‌بودن نمونه‌ها می‌باشد. ولی در شرح‌گذاری تصاویر هر نمونه بیش از یک برچسب دارد. بنابراین در بکارگیری الگوریتم‌های انتخاب نمونه‌های اولیه توجه به ویژگی چندبرچسبه بودن شرح‌گذاری امری ضروری است. در این رابطه بکارگیری معیار مناسب جهت محاسبه عدم‌شباخت بردار برچسب‌ها، از اهمیت ویژه‌ای برخوردار است. همچنین از آنجا که این روش‌ها مبتنی بر پیش‌بینی برچسب از نزدیکترین همسایه‌های نمونه کاندیدای حذف می‌باشد، بکارگیری رای‌گیری حداکثری سنتی از نزدیکترین همسایه‌ها، منجر به افت کارایی این روش‌ها می‌شود. در رویکرد MLENN-GKELM مطرح شده در این پایان‌نامه، در انتخاب نمونه‌های اولیه، از یک رابطه مناسب جهت محاسبه عدم‌شباخت دودویی دو بردار برچسب بهره برده شده است. همچنین به منظور پیش‌بینی برچسب از نزدیکترین همسایه‌ها، دو الگوریتم پیش‌بینی حریصانه و KELM استفاده شده است. بکارگیری این دو ایده در روش ENN پایه، منجر به بهبود عملکرد این روش در انتخاب نمونه‌های اولیه مناسب می‌شود. نتایج ارائه شده در سه بخش انتخاب نمونه‌های اولیه، شرح‌گذاری و تعادل مابین انتخاب نمونه‌های اولیه و شرح‌گذاری نشان از برتری رویکرد ارائه شده نسبت به روش‌های پایه‌ای دارد. این نتایج در جدول ۲-۵، جدول ۳-۵، جدول ۴-۵ و جدول ۵-۵ ارائه شده است.

ایده بهبود دقت شرح‌گذاری تصاویر بدون برچسب در این پایان‌نامه نیز بکارگیری روش‌های پیش‌بینی برچسب مناسب است. از این‌رو از دو روش یادگیری KNN و زن‌دار و KELM استفاده و دقت شرح‌گذاری تصاویر بدون برچسب آزمایش بهبود داده شده است. دقت شرح‌گذاری بدست آمده با سایر روش‌های شرح‌گذاری مجموعه تصاویر در مقیاس بزرگ مورد مقایسه قرار گرفته است. نتایج مقایسه‌های صورت گرفته در شکل ۴-۵، شکل ۵-۵ و شکل ۶-۵ و شکل ۷-۵ نشان داده شده است. نتایج تجربی بدست آمده نشان از صلاحیت عملکرد رویکردهای ارائه شده در کاربرد شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ دارد. در نهایت به منظور نشان دادن مقیاس‌پذیری رویکرد پیشنهادی، زمان شرح‌گذاری هر تصویر بدون برچسب و همچنین تجزیه و تحلیل مقیاس‌پذیری ارائه شد. نمودار زمان شرح‌گذاری هر تصویر بدون برچسب آزمایش در شکل ۸-۵ نشان داده شده است.

## ۶-۲-۶ مروری بر نوآوری‌ها

همانطور که در فصل اول بیان شد، تمرکز اصلی این پایان‌نامه بر روی حل چالش‌های مرتبط با مقیاس‌پذیری سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ می‌باشد. در این راستا، ایده اصلی، بکارگیری روش‌های انتخاب نمونه‌های اولیه در حل این چالش‌ها می‌باشد. اولین گام موثر این پایان‌نامه در جهت حل چالش زمان شرح‌گذاری، ارائه روشی ترکیبی به منظور یافتن زیر مجموعه نمونه‌های اولیه کمینه از مجموعه اصلی است. ایده اصلی در روش ترکیبی ارائه شده، در نظر گرفتن و حذف نمونه‌های پرت طبقه‌ای در روش CNN می‌باشد. این ایده تاکنون در دیگر پژوهش‌های مبتنی بر انتخاب نمونه‌های اولیه به آن پرداخته نشده است. گام موثر دوم این پایان‌نامه ارائه روش ENN چندبرچسبه در کاربرد شرح‌گذاری خودکار تصاویر می‌باشد. از طرفی در روش ارائه شده، اثربخشی بکارگیری روابط عدم شباهت دودویی متعدد بررسی شده است. همچنین با بکارگیری ایده روابط عدم شباهت تنظیم شده، روابط موجود تغییر یافته و در روش ارائه شده مورد استفاده قرار گرفته است.

گام موثر بعدی این پایاننامه، تغییر مرحله رای‌گیری و تصمیم‌گیری روش ENN با بکارگیری روش‌های پیش‌بینی برچسب موثر می‌باشد. در این مرحله به جای رای‌گیری حداکثری سنتی، بردار برچسبی از روی نزدیکترین همسایه‌ها ساخته می‌شود. در ساخت بردار برچسب مد نظر از دو روش پیش‌بینی حریصانه و KELM استفاده شده است.

پس از کاهش حجم مجموعه آموزش با روش PS چندبرچسبی با ایده‌های مختلف ارائه شده، تصاویر بدون برچسب با استفاده از روش‌های یادگیری شرح‌گذاری می‌شوند. به منظور بهبود دقت شرح‌گذاری تصاویر بدون برچسب، از دو روش یادگیری KNN وزن‌دار و KELM استفاده شده است.

پژوهش حاضر گامی در جهت حل چالش‌های سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ برمی‌دارد. رویکرد حل مسئله بکارگرفته شده در این پایاننامه با در نظر گرفتن زمان و مقیاس‌پذیری به عنوان پارامترهای اصلی سیستم، دقت سیستم را نیز در حد بالایی نگه می‌دارد. به بیان دقیق‌تر، این رویکرد، حد تعادلی مابین زمان و دقت سیستم در نظر خواهد گرفت.

### ۳-۶- پژوهش‌های پیشنهادی

به عنوان ایده‌های پژوهشی برای ادامه کار در حوزه‌های مورد بحث در این پایاننامه می‌توان به موارد زیر اشاره کرد. این ایده‌ها در دو قسمت شرح‌گذاری و انتخاب نمونه‌های اولیه موثر هستند. در ابتدا ایده‌هایی جهت ادامه کار در سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ و سپس ایده‌هایی در جهت ادامه کار در انتخاب نمونه‌های اولیه بیان می‌شوند.

۱- بکارگیری ایده‌ای در هر مرحله از سیستم‌های شرح‌گذاری خودکار مجموعه تصاویر در مقیاس بزرگ (بر اساس شکل ۲-۱) می‌تواند منجر به بهبود عملکرد شود. در مرحله ساخت بردار ویژگی سطح پایین از تصاویر، روش‌های متعددی می‌تواند بکار گرفته شود. ویژگی‌های سراسری، محلی و ترکیبی از ویژگی‌های سراسری و محلی متعدد منجر به بهبود دقت شرح‌گذاری خواهد شد.

۲- در این پایاننامه در ساخت بردار ویژگی هر تصویر صرفاً از تلفیق خطی سه دسته ویژگی سراسری استفاده شده است. این تلفیق ویژگی بر اساس پژوهش‌های صورت گرفته قبلی می‌باشد. در تلفیق ویژگی می‌توان از روابط تلفیق خطی و غیرخطی استفاده کرد و دقت

شرح‌گذاری را بهبود بخشد.

۳- در مرحله انتخاب و استخراج ویژگی به نسبت کارهای کمتری صورت گرفته است. به خصوص در این قسمت کمتر از روش‌های استخراج ویژگی بهره گرفته شده است. استفاده از الگوریتم‌های استخراج ویژگی می‌تواند ایده خوبی برای پژوهش در این حوزه باشد. در این زمینه بکارگیری روش‌های استخراج ویژگی خطی نمی‌تواند مناسب باشد. به منظور افزایش دقت شرح‌گذاری، استفاده از روش‌های استخراج ویژگی غیرخطی ایده مناسبی است.

۴- در بکارگیری مدل یادگیری و ساخت مدل پیش‌بینی، روش‌های نیمه‌نظرارتی به خوبی بکار گرفته شده و صلاحیت این روش‌ها به اثبات رسیده است. در محدود کارهایی از روش‌های انتخاب ویژگی در ترکیب با بکارگیری روش‌های یادگیری نیمه‌نظرارتی استفاده شده است. بکارگیری الگوریتم‌های مناسب در هر کدام از دو حوزه انتخاب و یا استخراج ویژگی، منجر به بهبود دقت در شرح‌گذاری خودکار تصاویر در مقیاس بزرگ خواهد شد. البته توجه به این نکته ضروری است که زمان شرح‌گذاری یک پارامتر بسیار مهم است که باید به خوبی به آن پرداخته شود.

۵- یکی از ایده‌هایی که در بهبود دقت شرح‌گذاری موثر خواهد بود، پالایش و رفع نویز تگ‌هاست. در این مرحله بکارگیری الگوریتمی مناسب می‌تواند منجر به بهبود دقت این سیستم‌ها شود.

۶- با توجه به پژوهش‌های صورت گرفته، توزیع‌پذیری محاسبات می‌تواند زمان شرح‌گذاری را در حد قابل قبولی نگه دارد. ولی عمدۀ پژوهش‌های صورت گرفته در این حوزه از روش‌های سنتی در یادگیری استفاده کرده‌اند. بکارگیری یک روش یادگیری نیمه‌نظرارتی در ترکیب

با یک روش انتخاب و یا استخراج ویژگی می‌تواند منجر به ارائه یک قاب کاری قدرتمند در حوزه شرح‌گذاری خودکار تصاویر با مقیاس بزرگ شود.

۷- در نهایت به منظور شرح‌گذاری در این پایان‌نامه از دو روش یادگیری مناسب استفاده شد.

در این مرحله نیز می‌توان روش‌های متعدد یادگیری را بکار برد. همچنین می‌توان از ایده ترکیب روش‌های یادگیری در جهت بهبود عملکرد پیش‌بینی استفاده کرد.

همه پیشنهادهای مطرح شده تابحال در جهت بهبود عملکرد شرح‌گذاری است. رویکرد حل

مسئله در این پایان‌نامه پیش‌پردازش داده‌ها با بکارگیری روش‌های PS است. اساس عملکرد انتخاب نمونه‌های اولیه نیز نرخ کاهش نمونه‌ها، دقت طبقه‌بندی و تعادل مابین نرخ کاهش و دقت طبقه‌بندی است. هر یک از این معیارها می‌تواند بهبود یابد. در جهت بهبود روش‌های PS در ادامه کار این پایان‌نامه، پژوهش‌های زیر پیشنهاد می‌شود.

۱- دو روش انتخاب نمونه اولیه بکار رفته در این پایان‌نامه، روش‌های پایه‌ای هستند. اساس

انتخاب این دو روش، سادگی در عین دقت بالا است. همانطور که در فصل سوم بیان شد، روش‌های PS متعددی مطرح شده است. اگرچه در کاربردهای مقیاس بزرگ استفاده از CNN برخی از این روش‌ها عملاً ممکن نیست؛ ولی بکارگیری روش‌های دیگری غیر ازENN و ENN می‌تواند عملکرد انتخاب نمونه‌های اولیه را بهبود دهد.

۲- هر یک از روش‌های PS نیز با ویژگی‌های بدست آمده برای نمونه‌ها کار می‌کنند. از این رو

بکارگیری روشی مناسب جهت ساخت بردار ویژگی، تلفیق دسته ویژگی‌ها و همچنین بکارگیری روش‌های انتخاب و استخراج ویژگی می‌تواند عملکرد انتخاب نمونه‌های اولیه را بهبود بخشد.

## واژه‌نامه فارسی به انگلیسی

انگلیسی	فارسی
Hyper plane	ابر صفحه
Edition	اصلاح‌سازی
Edited nearest neighbor	اصلاح‌سازی نزدیکترین همسایه
Multi-label edited nearest neighbor	اصلاح‌سازی نزدیکترین همسایه چندبرچسبه
2-fold cross validation	اعتبارسنجی متقابل دو بخشی
Incremental	افزایشی
Train	آموزش
Training set selection	انتخاب مجموعه آموزش
Instance selection	انتخاب نمونه
Prototype selection	انتخاب نمونه اولیه
Sparse feature selection	انتخاب ویژگی پراکنده
Back propagation	انتشار به عقب
Text-Based Image Retrieval	بازیابی تصویر مبتنی بر متن
Content-based image retrieval	بازیابی تصویر مبتنی بر محتوا
Texture	بافت
Supervised	بانظارت
Tuning free	بدون تنظیم
Online	برخط
Real time	بلادرنگ
Local frequency	بسامد محلی
Wrapper	بسته بندی
Boosting	بوستینگ
Parametric	پارامتری
Class outlier	پرت طبقه‌ای
Query	پرس و جو
Post processing	پس‌پردازش
Hybrid	ترکیبی
Text matching	تطابق متن
Mixed	تلغیقی

Adjusted	تنظیم شده
Condensation	تنکسازی
Condensed nearest neighbor	تنکسازی نزدیکترین همسایه
Class Outlier based Condensed Nearest Neighbor	تنکسازی نزدیکترین همسایه مبتنی بر حذف نمونه های پرت طبقه ای
Local descriptors	توصیفگرهای محلی
Fixed	ثابت
Multi-label	چندبرچسبه
Polynomial	چند جمله ای
Multi-class	چند کلاسه
Maximum mean discrepancy	حداکثر میانگین پراکندگی
Linear	خطی
Noisy	دارای اختلال
Auxiliary domain	دامنه کمکی
Cross-domain	دامنه متقابل
Decision trees	درختان تصمیم
Batch	دسته‌ای
Precision	دقت
Dudani	دودانی
Visual	دیداری
Majority voting	رأی گیری حداکثری
Fisher score	رتبه فیشر
Co-occurrence	رخداد مشترک
Digital	رقمی
Color	رنگ
Genetic	ژنتیک
Global	سراسری
Sigmoid	سیگموید
Artificial neural networks	شبکه های عصبی مصنوعی
Annotate	شرح‌گذاری
Image annotation	شرح‌گذاری تصویر
Automatic image annotation	شرح‌گذاری خودکار تصاویر
Shape	شکل

Regularization factor	ضریب تاثیر
Class	طبقه
Classifier	طبقه بند
Distance	عدم شباهت
Adjusted dissimilarity	عدم شباهت تنظیم شده
Performance	عملکرد
Non-parametric	غیر پارامتری
Recall	فراخوان
Meta data	فراداده
Filter	فیلتر
Framework	قابل کاری
Decremental	کاهشی
Bi-coded	کدشده دودویی
Correlogram	کورولوگرام
Bag of visual words	کیسه لغات دیداری
Gradient descend	گرادیان نزولی
Averaging stochastic gradient descent	گرادیان نزولی تصادفی
Pattern	الگو
Support vector machine	ماشین بردار پشتیبان
Extreme learning machine	ماشین یادگیری سریع
Kernel extreme learning machine	ماشین یادگیری سریع مبتنی بر هسته
Data set	مجموعه داده
Local	محلي
Repository	مخزن
Order dependent	مرتبه محور
Scale	مقیاس
Large scale	مقیاس بزرگ
Scalability	مقیاس پذیری
Wavelet texture	موجک بافت
Mean average precision	میانگین دقیق
Macro averaging	میانگین گیری ماکرو
Micro averaging	میانگین گیری میکرو

Weighted K-nearest neighbor	K-نزدیکترین همسایه وزن دار
Region	ناحیه
Trade-off rate	نرخ تعادل
Reduction rate	نرخ کاهش
Matrix norm	نرم ماتریسی
Map reduce	نگاشت کاهش
Hash	هش
Local hashing	هش محلی
Correlation	همبستگی
Hamming	همینگ
Keywords	کلید واژه‌ها
Linear weighting	وزن دهی خطی
Dual linear weighting	وزن دهی خطی دوگان
Inverse linear weighting	وزن دهی خطی معکوس
Logarithmic weighting	وزن دهی لگاریتمی
Rank weighting	وزن دهی مرتبه
Wilson	ویلسون
Graph laplacian-based semi-supervised learning	یادگیری نیمه‌نظری مبتنی بر گراف

## واژه‌نامه انگلیسی به فارسی

انگلیسی	فارسی
Adjusted	تنظیم شده
Adjusted dissimilarity	عدم شباهت تنظیم شده
Annotate	شرح گذاری
Artificial nueral networks	شبکه های عصبی
	مصنوعی
Automatic image annotation	شرح گذاری خودکار تصاویر
Auxiliary domain	دامنه کمکی
Averaging stochastic gradient descent	گرادیان نزولی تصادفی
Back propagation	انتشار به عقب
Bag of visual words	کیسه لغات دیداری
Batch	دسته‌ای
Bi-coded	کدشده دودویی
Boosting	بوستینگ
Class	طبقه
Class outiler	پرت طبقه‌ای
Class Outlier based Condensed Nearest Neighbor	تنکسازی نزدیکترین همسایه مبتنی بر حذف نمونه های پرت طبقه ای
Classifier	طبقه بند
Color	رنگ
Condensation	تنکسازی
Condensed nearest neighbor	تنکسازی نزدیکترین همسایه
Content-based image retrieval	بازیابی تصویر مبتنی بر محتوا
Co-occurance	رخداد مشترک
Correlation	همبستگی
Correlogram	کورولوگرام

Cross-domain	دامنه متقابل
Data set	مجموعه داده
Decision trees	درختان تصمیم
Decremental	کاهشی
Digital	رقمی
Distance	عدم شباهت
Dual linear weighting	وزن دهنده خطی دوگان
Dudani	دودانی
Edited nearest neighbor	اصلاح سازی نزدیکترین همسایه
Edition	اصلاح سازی
Extreme learning machine	ماشین یادگیری سریع
Filter	فیلتر
Fisher score	رتبه فیشر
Fixed	ثابت
Framework	قابل کاری
Genetic	ژنتیک
Global	سراسری
Gradient descend	گرادیان نزولی
Graph laplacian-based semi-supervised learning	یادگیری نیمه ناظارتی مبتنی بر گراف
Hamming	همینگ
Hash	هش
Hybrid	ترکیبی
Hyper plane	ابر صفحه
Image annotation	شرح گذاری تصویر
Incremental	افزایشی
Instance selection	انتخاب نمونه
Inverse linear weighting	وزن دهنده خطی معکوس
Kernel extreme learning machine	ماشین یادگیری سریع مبتنی بر هسته
Keywords	کلید واژه ها
Large scale	مقیاس بزرگ

Linear	خطی
Linear weighting	وزن دهی خطی
Local	محلي
Local descriptors	توصیفگرهای محلی
Local frequency	بسامد محلی
Local hashing	هش محلی
Logarithmic weighting	وزن دهی لگاریتمی
Macro averaging	میانگین گیری ماکرو
Majority voting	رای گیری حداکثری
Map reduce	نگاشت کاهش
Matrix norm	نرم ماتریسی
Maximum mean discrepancy	حداکثر میانگین
	پراکندگی
Mean average precision	میانگین دقต
	فراداده
Meta data	
Micro averaging	میانگین گیری میکرو
Mixed	تلفیقی
Multi-class	چند کلاسه
Multi-label	چندبر چسبه
Multi-label edited nearest neighbor	اصلاح سازی نزدیکترین همسایه چندبر چسبه
Noisy	دارای اختلال
Non-parametric	غیر پارامتری
Online	برخط
Order dependent	مرتبه محور
Parametric	پارامتری
Pattern	الگو
Performance	عملکرد
Polynomial	چند جمله ای
Post processing	پس پردازش
Precision	دقت
Prototype selection	انتخاب نمونه اولیه

Query	پرس و جو
Rank weighting	وزن دهی مرتبه
Real time	بلادرنگ
Recall	فراخوان
Reduction rate	نرخ کاهش
Region	ناحیه
Regularization factor	ضریب تاثیر
Repository	مخزن
Scalability	مقیاس پذیری
Scale	مقیاس
Shape	شكل
Sigmoid	سیگموید
Sparse feature selection	انتخاب ویژگی پراکنده
Supervised	بانظارت
Support vector machine	ماشین بردار پشتیبان
Text matching	تطابق متن
Text-Based Image Retrieval	بازیابی تصویر مبتنی بر متن
Texture	بافت
Trade-off rate	نرخ تعادل
Train	آموزش
Training set selection	انتخاب مجموعه آموزش
Tuning free	بدون تنظیم
Visual	دیداری
Wavelet texture	موجک بافت
Weighted K-nearest neighbor	K-نزدیکترین همسایه وزن دار
Wilson	ویلسون
Wrapper	بسته بندی

## مراجع

- [1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, “A survey of content-based image retrieval with high-level semantics,” *Pattern Recognit.*, vol. 40, no. 1, pp. 262–282, 2007.
- [2] X. Ke, S. Li, and G. Chen, “Real web community based automatic image annotation,” *Comput. Electr. Eng.*, vol. 39, no. 3, pp. 945–956, 2013.
- [3] F. Wang, “A survey on automatic image annotation and trends of the new age,” in *Procedia Engineering*, pp. 434 – 438, 2011.
- [4] Y. Yang, F. Wu, F. Nie, H. T. Shen, Y. Zhuang, and A. G. Hauptmann, “Web and Personal Image Annotation by Mining Label Correlation With Relaxed Visual Graph Embedding,” *IEEE Trans. image Process.*, vol. 21, no. 3, pp. 1339–1351, 2012.
- [5] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, “NUS-WIDE : A Real-World Web Image Database from National University of Singapore,” in *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009, pp. 0–8.
- [6] E. Alpaydin, *introduction to machine learning*, 2nd ed. London, England: The MIT Press, 2010.
- [7] Y.-G. Jiang, Q. Dai, J. Wang, C.-W. Ngo, X. Xue, and S.-F. Chang, “Fast semantic diffusion for large-scale context-based image and video annotation,” *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 80–91, 2012.
- [8] D. Liu and X. H. H. Zhang, “Content-based tag processing for Internet social images,” *Multimed Tools Appl*, vol. 51, no. November 2010, pp. 723–738, 2011.
- [9] M. A. Z. Chahooki and N. M. Charkari, “Shape classification by manifold learning in multiple observation spaces,” *Inf. Sci. (Ny)*., vol. 262, pp. 46–61, 2014.

- [10] R. Li, J. Lu, Y. Zhang, Z. Lu, and W. Xu, “A Framework of Large-scale and Real-time Image Annotation System,” in International Joint Conference on Artificial Intelligence, 2009, pp. 576–579.
- [11] ح. کارگر شورکی و م. ع. زارع چاهوکی, “شرح گذاری خودکار مجموعه تصاویر در مقیاس بزرگ مبتنی بر بهبود روش تنکسازی نزدیکترین همسایه با حذف نمونه‌های پرت طبقه‌ای ”, دومین کنفرانس بین المللی پردازش تصویر و بازناسی الگو ایران، دانشگاه گیلان. ۱۳۹۳.
- [12] F. Charte, A. J. Rivera, M. J. del Jesus, and F. Herrera, “MLeNN: A First Approach to Heuristic Multilabel Undersampling,” in Intelligent Data Engineering and Automated Learning--IDEAL 2014, 2014, pp. 1–9.
- [13] H. Kargar-Shooroki and M. A. Zare Chahooki, “Selection of effective training instances for scalable automatic image annotation,” Multimed. Tools Appl.
- [14] H. Kargar-Shooroki and M. A. Zare Chahooki, “GWML-ENN: A two-phase greedy-weighted framework for selection of effective labeled instances in scalable automatic image annotation,” Mach. Vis. Appl.
- [15] H. Kargar-Shooroki, M. A. Zare Chahooki, and S. Javanmardi, “MLENN-KELM: a prototype selection based kernel extreme learning machine approach for large-scale automatic image annotation,” Adv. Comput. Sci. an Int. J., vol. 4, no. 5, pp. 95–100, 2015.
- [16] ح. کارگر شورکی و م. ع. زارع چاهوکی, “روشی مبتنی بر ماشین یادگیری سریع با هسته غیرخطی برای انتخاب نمونه‌های اولیه در یادگیری چندبرچسبی مقیاس بزرگ ”, ماشین بینایی و پردازش تصویر.
- [17] C. Shi, Q. Ruan, and G. An, “Sparse feature selection based on graph Laplacian for web image annotation,” Image Vis. Comput., vol. 32, no. 3, pp. 189–201, 2014.

- [18] Z. Ma, F. Nie, Y. Yang, J. R. R. Uijlings, N. Sebe, and S. Member, “Web Image Annotation Via Subspace-Sparsity Collaborated Feature Selection,” IEEE Trans. Multimed., vol. 14, no. 4, pp. 1021–1030, 2012.
- [19] D. Zhang and W.-J. Li, “Large-scale supervised multimodal hashing with semantic correlation maximization,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2014.
- [20] M. Kan, D. Xu, S. Shan, and X. Chen, “Semisupervised Hashing via Kernel Hyperplane Learning for Scalable Image Search,” Circuits Syst. Video Technol. IEEE Trans., vol. 24, no. 4, pp. 704–713, 2014.
- [21] C. Deng, X. Liu, Y. Mu, and J. Li, “Large-scale multi-task image labeling with adaptive relevance discovery and feature hashing,” Signal Processing, vol. 112, pp. 137–145, 2015.
- [22] X. Chen, Y. Mu, S. Yan, and T.-S. Chua, “Efficient large-scale image annotation by probabilistic collaborative multi-label propagation,” in Proceedings of the international conference on Multimedia, 2010, pp. 35–44.
- [23] S. H. Amiri and M. Jamzad, “Efficient Multi-modal Fusion on Supergraph for Scalable Image Annotation,” Pattern Recognit., 2015.
- [24] K.-S. Hamid and Z.-C. Mohammad-Ali, “Large-Scale Automatic Image Annotation Based on Kernelized kNN,” in 8th Iranian Data Mining Conference (In Persian), 2014.
- [25] W.-Y. Lee, L.-C. Hsieh, G.-L. Wu, and W. Hsu, “Graph-based semi-supervised learning with multi-modality propagation for large-scale image datasets,” J. Vis. Commun. Image Represent., vol. 24, no. 3, pp. 295–302, 2013.

- [26] Y. Yuan, F. Wu, J. Shao, and Y. Zhuang, “Image annotation by semi-supervised cross-domain learning with group sparsity,” *J. Vis. Commun. Image Represent.*, vol. 24, no. 2, pp. 95–102, 2013.
- [27] X. Zhu, “Semi-Supervised Learning Literature Survey,” 2005.
- [28] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, and N. E. C. L. America, “Large-scale Image Classification : Fast Feature Extraction and SVM Training,” in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Conference on., 2010.
- [29] J. Liu, R. Hu, M. Wang, Y. Wang, and E. Y. Chang, “Web-scale Image Annotation,” in *Advances in Multimedia Information Processing - PCM 2008*, 2008, pp. 663–674.
- [30] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer, 2015.
- [31] S. Garcia, J. Derrac, J. R. Cano, and F. Herrera, “Prototype Selection for Nearest Neighbor Classification : Taxonomy and Empirical Study,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 34, no. 3, pp. 417–435, 2012.
- [32] P. E. Hart, “The Condensed Nearest Neighbor Rule,” *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 515–516, 1968.
- [33] V. S. Devi and M. N. Murty, “An incremental prototype set building technique,” *Pattern Recognit.*, vol. 35, no. 2, pp. 505–513, 2002.
- [34] G. Gates, “The reduced nearest neighbor rule (Corresp.),” *Inf. Theory, IEEE Trans.*, vol. 18, no. 3, pp. 431–433, 1972.
- [35] F. Angiulli, “Fast Nearest Neighbor Condensation for Large Data Sets Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 11, pp. 1450–1464, 2007.
- [36] D. L. Wilson, “Asymptotic properties of nearest neighbor rules using edited data,” *Syst. Man Cybern. IEEE Trans.*, no. 3, pp. 408–421, 1972.

- [37] I. Tomek, “An experiment with the edited nearest-neighbor rule,” *IEEE Trans. Syst. Man. Cybern.*, no. 6, pp. 448–452, 1976.
- [38] P. A. Devijver, “On the editing rate of the multiedit algorithm,” *Pattern Recognit. Lett.*, vol. 4, no. 1, pp. 9–12, 1986.
- [39] J. S. Sánchez, F. Pla, and F. J. Ferri, “Prototype selection for the nearest neighbour rule through proximity graphs,” *Pattern Recognit. Lett.*, vol. 18, no. 6, pp. 507–513, 1997.
- [40] D. R. Wilson and T. R. Martinez, “Reduction techniques for instance-based learning algorithms,” *Mach. Learn.*, vol. 38, no. 3, pp. 257–286, 2000.
- [41] H. Brighton and C. Mellish, “Advances in instance selection for instance-based learning algorithms,” *Data Min. Knowl. Discov.*, vol. 6, no. 2, pp. 153–172, 2002.
- [42] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991.
- [43] J. C. Riquelme, J. S. Aguilar-Ruiz, and M. Toro, “Finding representative patterns with ordered projections,” *Pattern Recognit.*, vol. 36, no. 4, pp. 1009–1018, 2003.
- [44] S. García, J. R. Cano, and F. Herrera, “A memetic algorithm for evolutionary prototype selection: A scaling up approach,” *Pattern Recognit.*, vol. 41, no. 8, pp. 2693–2709, 2008.
- [45] L. I. Kuncheva, “Editing for the k-nearest neighbors rule by a genetic algorithm,” *Pattern Recognit. Lett.*, vol. 16, no. 8, pp. 809–814, 1995.
- [46] N. García-Pedrajas, J. A. R. Del Castillo, and D. Ortiz-Boyer, “A cooperative coevolutionary algorithm for instance selection for instance-based learning,” *Mach. Learn.*, vol. 78, no. 3, pp. 381–420, 2010.

- [47] D. B. Skalak, “Prototype and feature selection by sampling and random mutation hill climbing algorithms,” in Proceedings of the eleventh international conference on machine learning, 1994, pp. 293–301.
- [48] V. S. Devi and M. N. Murty, “An incremental prototype set building technique,” *Pattern Recognit.*, vol. 35, no. 2, pp. 505–513, 2002.
- [49] N. M. Hewahi and M. K. Saad, “Class Outliers Mining : Distance-Based Approach,” *Int. J. Intell. Syst. Technol.*, vol. 2, no. 9, pp. 2803–2816, 2007.
- [50] S.-S. Choi, S.-H. Cha, and C. C. Tappert, “A survey of binary similarity and distance measures,” *J. Syst. Cybern. Informatics*, vol. 8, no. 1, pp. 43–48, 2010.
- [51] A. Makadia, V. Pavlovic, and S. Kumar, “Baselines for image annotation,” *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 88–105, 2010.
- [52] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *Syst. Man, Cybern. Part B Cybern. IEEE Trans.*, vol. 42, no. 2, pp. 513–529, 2012.
- [53] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, vol. 2, pp. 985–990.
- [54] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [55] W. Zong and G.-B. Huang, “Face recognition based on extreme learning machine,” *Neurocomputing*, vol. 74, no. 16, pp. 2541–2551, 2011.
- [56] B. Lu, G. Wang, Y. Yuan, and D. Han, “Semantic concept detection for video based on extreme learning machine,” *Neurocomputing*, vol. 102, pp. 176–183, 2013.

- [57] J. Kim, H. Shin, Y. Lee, and M. Lee, “Algorithm for classifying arrhythmia using Extreme Learning Machine and principal component analysis,” in Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, 2007, pp. 3257–3260.
- [58] G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” Neural Networks, IEEE Trans., vol. 17, no. 4, pp. 879–892, 2006.
- [59] G. Feng, G.-B. Huang, Q. Lin, and R. Gay, “Error minimized extreme learning machine with growth of hidden nodes and incremental learning,” Neural Networks, IEEE Trans., vol. 20, no. 8, pp. 1352–1357, 2009.
- [60] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks,” Neural Networks, IEEE Trans., vol. 17, no. 6, pp. 1411–1423, 2006.
- [61] Z. Lin, J. Cao, T. Chen, Y. Jin, Z.-L. Sun, and A. Lendasse, “Extreme Learning Machine on High Dimensional and Large Data Applications,” Math. Probl. Eng., vol. 501, p. 624903, 2015.
- [62] T.-L. Pao, Y.-T. Chen, J.-H. Yeh, Y.-M. Cheng, and Y.-Y. Lin, “A comparative study of different weighting schemes on KNN-based emotion recognition in Mandarin speech,” in Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, Springer, 2007, pp. 997–1005.
- [63] J. Gou, L. Du, Y. Zhang, and T. Xiong, “A new distance-weighted k-nearest neighbor classifier,” J. Inf. Comput. Sci, vol. 9, pp. 1429–1436, 2012.
- [64] T. Bailey and A. K. Jain, “A Note on Distance-Weighted k -Nearest Neighbor Rules,” IEEE Trans. Syst. Man. Cybern., no. 4, pp. 311–313, 1978.

- [65] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip, and others, “Top 10 algorithms in data mining,” *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [66] S. A. Dudani, “The distance-weighted k-nearest-neighbor rule,” *Syst. Man Cybern. IEEE Trans.*, no. 4, pp. 325–327, 1976.
- [67] M.-L. Zhang and Z.-H. Zhou, “A review on multi-label learning algorithms,” *Knowl. Data Eng. IEEE Trans.*, vol. 26, no. 8, pp. 1819–1837, 2014.
- [68] H. Bannour, L. Hlaoua, and B. Ayeb, “Survey of the Adequate Descriptor for Content-Based Image Retrieval on the Web : Global versus Local Features,” Tunisia, 2009.
- [69] A. Özgür, L. Özgür, and T. Güngör, “Text categorization with class-based and corpus-based keyword selection,” in *Computer and Information Sciences-ISCIS* 2005, Springer, 2005, pp. 606–615.
- [70] W. Deng, Q. Zheng, and K. Zhang, “Reduced kernel extreme learning machine,” in *Proceedings of the 8th International Conference on Computer Recognition Systems CORES* 2013, 2013, pp. 63–69.
- [71] F. Wang and C. Zhang, “Label propagation through linear neighborhoods,” *Knowl. Data Eng. IEEE Trans.*, vol. 20, no. 1, pp. 55–67, 2008.
- [72] A. Subramanya and J. A. Bilmes, “Entropic graph regularization in non-parametric semi-supervised classification,” in *Advances in Neural Information Processing Systems*, 2009, pp. 1803–1811.
- [73] J. Tang, R. Hong, S. Yan, and T. Chua, “Image Annotation by k NN-Sparse Graph-based Label Propagation over Noisily-Tagged Web Images,” vol. 1, no. 1, pp. 111–126, 2010.

[74] J. Huang, H. Liu, J. Shen, and S. Yan, “Towards efficient sparse coding for scalable image annotation,” in Proceedings of the 21st ACM international conference on Multimedia, 2013, pp. 947–956.

**Abstract**— Automatic image annotation means employing learning models to describe the visual content of digital images by using text descriptors. Annotation models deal with new challenges due to the large number of images and annotation words in image sets. The most important of these challenges are system scalability, annotation time and annotation accuracy. In this thesis, to address large-scale image annotation challenges, prototype selection idea is utilized.

In the use of prototype selection idea in large scale automatic image annotation, two new approaches are proposed. In the first approach, class outlier detection idea is employed in CNN method and an extended version of this method is proposed. Automatic image annotation is a multi-label classification problem. In the second approach, to improve performance and compliance of prototype selection methods in multi-label domains, a new method based on ENN is proposed. In the proposed MLENN-GKELM method, decision about removing each instance is made in two steps. At first, label vector of candidate removal instance is predicted from its nearest neighbors by using two Greedy and KELM prediction algorithms. Then, decision about removing each instance is made based on multi-label dissimilarity of two predicted and real label vectors. In this step, the effectiveness of using various binary dissimilarity measures is provided. These ideas are two main contributions of the second approach. Prototype selection performance in ENN is improved by using these ideas. In the annotation procedure, two weighted KNN and KELM algorithms are employed and annotation performance is improved.

The efficiency and effectiveness of the proposed approaches in selecting effective instances and improving annotation performance has been shown with extensive experiments on large-scale image sets. In this regard, NUS-WIDE-Object, NUS-WIDE-Scene, NUS-WIDE-Lite and NUS-WIDE image sets have been used. Experimental results in reduction rate, annotation performance and trade-off between reduction and annotation are presented. Reduction rates of proposed approaches in

these image sets are 74.15%, 60.68%, 79.09% and 90.43% respectively. As well as annotation performance in terms of macro F1 are 36.31%, 33.24%, 25.63% and 22.72% respectively. Finally, trade-off performance between reduction and annotation in these images sets are 26.92%, 20.17%, 20.27% and 20.55% respectively. Experimental results revealed that large-scale image annotation challenges are enhanced in the proposed approaches.

**Keyword:** Automatic image annotation, Large-scale image sets, Prototype selection, Scalability and annotation time.

**Yazd University**  
**Faculty of Electrical and Computer Engineering**

**A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Master Degree in Computer Engineering**

**Title**

**Large scale automatic image annotation  
based on prototype selection**

**Supervisor**

**Dr.M.A.Zare Chahooki**

**Advisor**

**Dr.M.Rezaeian**

**By**

**Hamid Kargar Shooroki**

**Winter 2015**