دانشگاه فردوسی مشهد
Ferdowsi University
of Mashhad

دانشکده مهندسی
گروه مهندسی کامپیوتر
مقطع کارشناسی

عنوان

SVM and Decision Tree

نگارش
امیررضا خادم پیر
حانیه قوی پنجه

استاد درس
دکتر احسان فضل ارثی

آذر ۱۴۰۴

فهرست مطالب

فاز اول - **Data Exploration and Profiling** ۲
فاز دوم - **Feature Extraction** ۵
فاز سوم - **Train** ۶
آموزش مدل Baseline بدون تیونینگ ۶
تیونینگ مدل‌ها ۷
فاز چهارم - **Evaluation** ۷
معیارهای ارزیابی ۸
نتایج ارزیابی مدل‌ها ۸
فاز پنجم - **Image Segmentation, Unimportant Region Removal** ۱۰
مرحله Segmentation تولید سوپرپیکسل‌ها با الگوریتم SLIC ۱۱
خوشه‌بندی دو مرحله‌ای ۱۱
محاسبه اهمیت نواحی با مدل SVM ۱۲
انتخاب و ماسک نواحی غیرمهم ۱۳
فاز ششم - **Train2** ۱۳
فاز هفتم - **Evaluation2** ۱۴
نتایج ارزیابی مدل‌ها ۱۴
تحلیل افت دقت پس از حذف خوشه‌ها ۱۵
فاز هشتم - نمره اضافه ۱۵

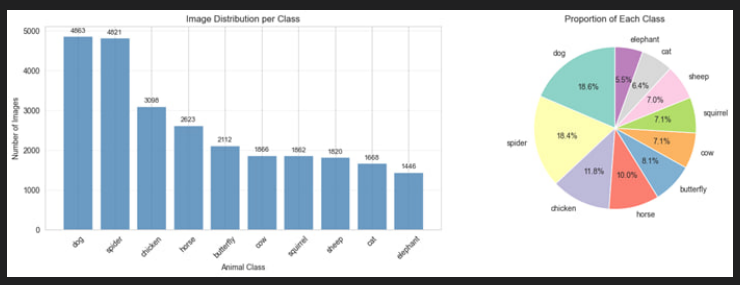
فاز اول - Data Exploration and Profiling

همینطور که دیده میشود عدم تعادل بین کلاسها وجود داره(کلاس dog با بیشترین و elephant با کمترین) این عدم تعادل قابل توجه است و می‌تواند مدل های طبقه بندی رو به سمت کلاس‌های غالب bias بکنه. مدل‌های پایه مثل SVM و Tree Decision احتمالاً برای کلاس‌های بزرگ مثل dog بایاس خواهند داشت.

```
Distribution of images per class:
    Class  Count
      dog   4863
   spider   4821
  chicken   3098
    horse   2623
butterfly   2112
      cow   1866
 squirrel   1862
    sheep   1820
      cat   1668
 elephant   1446

Total images: 26179
Average images per class: 2617.90
Median images per class: 1989.00
```

شکل ۱: توزیع کلاس های Dataset



شکل ۲: توزیع کلاس های Dataset

بررسی ویژگی‌های سطح فایل و ساختار تصاویر:
ابعاد تصاویر: تصاویر تنوع بالایی در اندازه دارند مثلا عرض تصاویر از حدود ۶۰ پیکسل تا بیش از ۶۰۰۰ پیکسل متغیر است. هیچ یکنواختی کاملی وجود ندارد، که پیش‌پردازش(مثل Resize) رو ضروری میکنه.
حالت رنگ: توزیع حالت رنگ تصاویر بر اساس نتایج:

- RGB : تقریباً اکثریت غالب
- RGBA : تعداد کمی (۴۹ تصویر)
- L (سیاه‌سفید) : فقط ۱ تصویر
- CMYK : فقط ۱ تصویر

دیتاست تقریباً یکپارچه از نظر حالت رنگ است (اکثراً (RGB موارد غیر RGB بسیار کم‌اند و می‌توان به راحتی آن‌ها را به RGB تبدیل کرد.

```
========================================
IMAGE DIMENSION STATISTICS
========================================
Width  - Mean: 320.04, Std: 196.94
         Min: 60, Max: 6720

Height - Mean: 252.63, Std: 148.40
         Min: 57, Max: 6000

Aspect Ratio - Mean: 1.31
               Range: [0.29, 4.23]
```

شکل ۳: نتایج تحلیل ویژگی های سطح فایل

بررسی ویژگی‌های محتوایی تصاویر:

۱. روشنایی و کنتراست
بررسی تصویری (خروجی‌های نموداری) نشان می‌دهد:

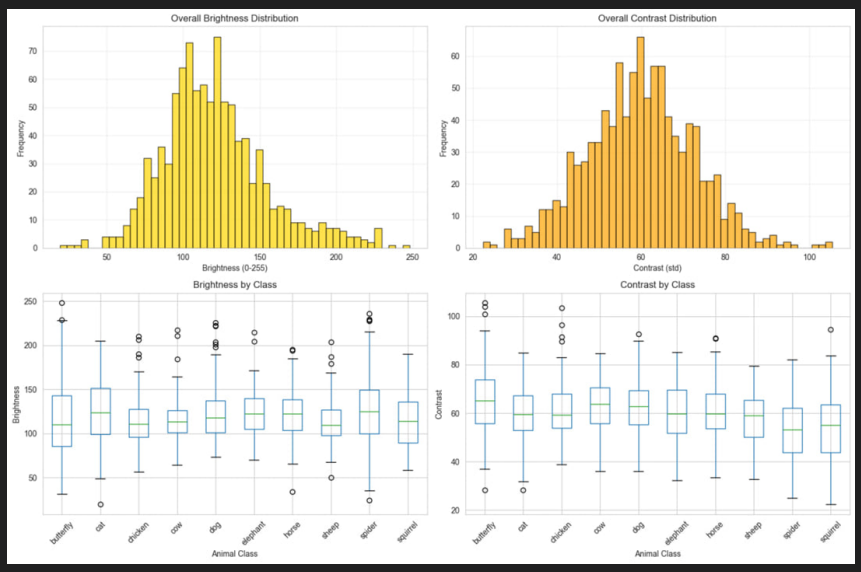
- پراکندگی روشنایی در برخی کلاس‌ها بیشتر است
- کلاس‌هایی که در محیط‌های طبیعی ثبت شده‌اند مانند sheep و horse نور متغیرتری دارند
- کلاس‌هایی مثل dog یا cat توزیع یکنواخت‌تری از روشنایی دارند

این ناهمگونی احتمالاً روی ویژگی‌های بافتی و شدت‌پایه مثل LBP یا GLCM اثر می‌گذارد و بهتر است در فاز پیش‌پردازش، نرمال‌سازی روشنایی مدنظر قرار گیرد.

Visual Properties Summary by Class:

	brightness		contrast		saturation	
	mean	std	mean	std	mean	std
class						
butterfly	120.41	49.88	65.84	14.50	0.45	0.22
cat	126.27	36.98	60.13	11.56	0.26	0.16
chicken	113.34	29.32	61.39	12.54	0.33	0.17
cow	115.17	25.18	62.73	10.11	0.35	0.14
dog	123.99	33.92	62.85	11.16	0.30	0.14
elephant	123.18	26.03	60.31	11.51	0.30	0.13
horse	121.75	28.03	60.73	12.50	0.32	0.13
sheep	114.27	25.54	57.69	10.73	0.35	0.13
spider	128.33	42.88	53.21	12.71	0.35	0.21
squirrel	116.49	31.52	54.26	14.46	0.33	0.15

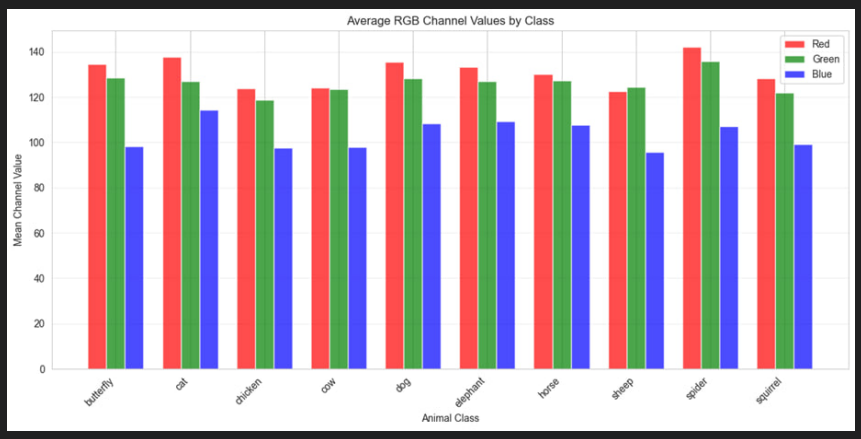
شکل ۴: نتایج تحلیل روشنایی و کنتراست



شکل ۵: نمودار هیستوگرام و باکس پلات روشنایی و کنتراست

۲. توزیع رنگ ها
در بررسی توزیع فراوانی رنگ مشخص شد که:

- برخی کلاس‌ها رنگ‌های غالب مشخص دارند
  - سبز و قهوه‌ای برای حیوانات محیط باز، cow، horse sheep به‌دلیل پس‌زمینه چمن و مزرعه مثل
  - طیف خاکستری و قهوه‌ای برای squirrel یا elephant
  - رنگ‌های روشن‌تر و فاقد پس‌زمینه طبیعی برای butterfly
- این ویژگی‌ها نشان می‌دهند که ویژگی‌های رنگی می‌توانند نقش مهمی در جداکردن کلاس‌ها داشته باشند.



شکل ۶: توزیع RGB کلاس ها

نتیجه گیری:

۱. دیتاست نامتوازن است و این موضوع روی مدل‌های پایه اثر منفی خواهد داشت.

۲. اندازه تصاویر یکسان نیست و نیاز به Resize استاندارد وجود دارد.

۳. حالت رنگ تقریباً یکپارچه است (RGB) ، ولی چند نمونه استثنا باید تبدیل شوند

۴. روشنایی، کنتراست و پس‌زمینه‌ها یکنواخت نیستند؛ این یعنی ویژگی‌های intensity-base ممکن است متغیر باشند

۵. رنگ‌های غالب بین کلاس‌ها قابل تشخیص هستند و می‌توانند برای تفکیک استفاده شوند.

Detailed Statistics per Class:

Class	Count	Avg Width	Avg Height	Brightness	Contrast	Variance	Edge Density
butterfly	2112	364.55	280.42	120.41	65.84	3445.22	68.43
cat	1668	473.16	364.57	126.27	60.13	3573.88	59.29
chicken	3098	288.30	239.36	113.34	61.39	3022.84	107.39
cow	1866	294.18	212.75	115.17	62.73	3522.49	91.87
dog	4863	283.57	241.09	123.99	62.85	3407.40	78.26
elephant	1446	377.62	274.92	123.18	60.31	3185.67	84.42
horse	2623	287.95	233.47	121.75	60.73	4170.35	93.66
sheep	1820	363.05	263.37	114.27	57.69	3119.16	89.66
spider	4821	311.14	247.83	128.33	53.21	2457.96	79.94
squirrel	1862	287.84	224.65	116.49	54.26	2962.70	86.98

شکل ۷: نتایج کلی

فاز دوم - Feature Extraction

ابتدا تمام تصاویر به اندازه ثابت 128×128 تبدیل شدند تا داده‌ها از نظر ابعاد یکنواخت شده و مقایسه ویژگی‌ها میان نمونه‌ها قابل‌اعتماد باشد. پس از این مرحله، مجموعه‌ای از ویژگی‌ها طراحی و استخراج شد که هر کدام بخشی از اطلاعات رنگی، بافتی، هندسی یا ساختاری تصویر را ثبت می‌کنند.

به طور کلی، این ویژگی‌ها به چند گروه اصلی تقسیم می‌شوند. ابتدا ویژگی‌های هندسی مانند نسبت ابعاد تصویر استخراج شد تا اطلاعات کلی درباره ساختار اولیه تصویر در اختیار مدل قرار گیرد. سپس ویژگی‌های رنگی در دو فضای RGB و HSV محاسبه شدند؛ دلیل انتخاب این دو فضا آن است که RGB اطلاعات رنگ خام را ثبت می‌کند، در حالی که HSV بیشتر با ویژگی‌های ادراکی انسان هماهنگ بوده و تحلیل شدت روشنایی و اشباع را آسان‌تر می‌سازد.

در ادامه، ویژگی‌های شدت و روشنایی شامل میانگین، کنتراست ساده، چولگی، کشیدگی و آنتروپی استخراج شد که توزیع نور و شدت پیکسل‌ها را توصیف می‌کنند و برای تفکیک کلاس‌هایی که ظاهر یکنواخت‌تر یا متغیرتر دارند مفید هستند. برای تحلیل بافت نیز از ویژگی‌های GLCM و LBP استفاده شد؛ GLCM الگوهای هم‌وقوعی پیکسل‌ها را ثبت می‌کند و شاخص‌هایی مانند انرژی، همگنی و همبستگی را ارائه می‌دهد، در حالی که LBP یک توصیف‌گر محلی قدرتمند برای الگوهای بافتی ریز است.

در کنار این‌ها، ویژگی‌های لبه با استفاده از فیلترهای گرادیانی محاسبه شد تا شدت تغییرات ساختاری و تراکم لبه‌ها مشخص شود. همچنین ویژگی‌های شکل بر اساس بزرگ‌ترین کانتور استخراج گردید تا اطلاعاتی درباره مساحت، محیط و میزان فشردگی ناحیه اصلی موجود در تصویر به دست آید. در نهایت، توصیف‌گر گرادیانی HOG نیز استخراج شد که الگوهای جهت‌گیری لبه‌ها را ثبت کرده و در تشخیص ساختارهای کلی حیوانات بسیار مؤثر است.

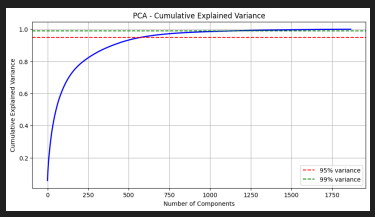
تمام این ویژگی‌ها در نهایت به صورت یکه عددی تجمیع شده و به همراه نام فایل، کلاس و برچسب عددی در فایل feature2.csv ذخیره شدند. این داده ساختاریافته ورودی اصلی فاز سوم است و برای مدل‌سازی و تیونینگ مناسبه.

فاز سوم - Train

آماده‌سازی داده‌ها شامل نرمال‌سازی ویژگی‌ها بود؛ زیرا داده‌های استخراج‌شده از جنس‌های مختلف (بافت، رنگ، شکل، شدت نور و ...) دارای مقیاس‌های کاملاً متفاوت بودند و این اختلاف می‌توانست بر رفتار مدل‌ها، خصوصاً SVM، اثر منفی بگذارد. پس از استانداردسازی، داده به مجموعه‌های Train و Test تقسیم شد با نسبت ۸۰/۲۰.

کاهش‌بعد با استفاده از PCA

به دلیل آنکه بردار ویژگی‌ها شامل حدود ۱۸۰۰ مقدار عددی بودند، اجرای مستقیمSVM خصوصاً در هنگام تیونینگ، بسیار زمان‌بر می‌شد. برای حل این مسئله، از PCA استفاده شد. این کار باعث شد که هم سرعت آموزش مدل‌ها و هم سرعت جست‌وجوی بهینه‌سازی به‌طور قابل توجهی افزایش یابد، بدون آنکه دقت مدل‌ها دچار افت محسوسی شود.



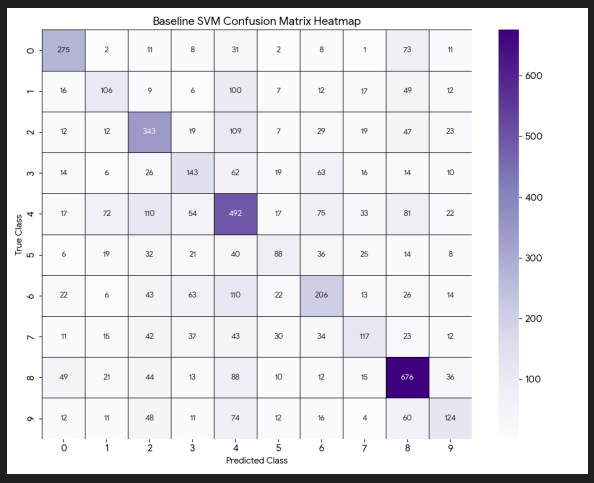
شکل ۸: نمودار واریانس

آموزش مدل Baseline بدون تیونینگ

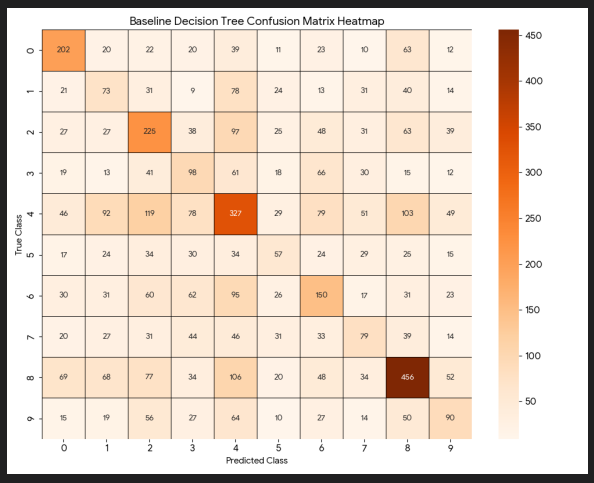
آ. SVM بدون تیونینگ
با استفاده از کرنل خطی و تنظیمات پیش‌فرض.

ب. درخت تصمیم بدون تیونینگ
با عمق ۲۰ و معیار پیش‌فرض Gini

نتایج:
SVM عملکرد متوسط و نسبتاً پایدار نشان داد، اما با داده‌های نامتوازن برای برخی کلاس‌ها بایاس داشت. این نتایج نشان داد که تیونینگ باید انجام بشه.



شکل ۹: Confusion Matrix بدون تیونینگ SVM



شکل ۱۰: Confusion Matrix بدون تیونینگ DT

accuracy	0.61

شکل ۱۱: دقت SVM

accuracy	0.31

شکل ۱۲: دقت DT

تیونینگ مدل‌ها

تیونینگ هایپرپارامترها با استفاده از روش RandomizedSearchCV انجام شد. دلیل انتخاب این روش آن بود که فضای جست‌وجوی پارامترها، به‌ویژه برای SVM شامل پارامترهایی مانند C، gamma و نوع کرنل، بسیار گسترده بود و استفاده از GridSearch به‌معنای صرف زمان چندین برابر و هزینهٔ محاسباتی شدیدتر بود. روش تصادفی به‌دلیل آن که امکان آزمون سریع بخش مؤثری از فضای جست‌وجو را فراهم می‌کند، توانست با تعداد محدودی تکرار به ترکیب‌هایی برسد که عملکرد نزدیک به بهینه ارائه می‌دادند و در عین حال سرعت کل فرآیند را حفظ می‌کردند.

در تیونینگ SVM، ترکیب بهینه شامل کرنل RBF، مقدار متوسط برای C و مقدار تنظیم‌شده برای gamma بود. این ترکیب از آن جهت بهترین نتیجه را داد که کرنل RBF توانایی مناسبی در مدل‌سازی مرزهای غیرخطی در فضای ویژگی داشت، مقدار C باعث حفظ تعادل میان بایاس و واریانس شد و مقدار gamma نیز حساسیت مدل به جزئیات و نویز را کنترل کرد.

در مورد درخت تصمیم، روند جست‌وجو نشان داد که نسخه‌های بیش‌ازحد عمیق با کم‌ترین محدودیت، به‌سرعت دچار بیش‌برازش می‌شوند. به همین دلیل بهترین ترکیب شامل محدود کردن عمق، افزایش min_samples_split و بالا بردن min_samples_leaf شد. این سه محدودیت باعث شدند درخت به‌جای دنبال‌کردن نوسانات تصادفی داده، الگوهای پایدارتر را یاد بگیرد و بخش‌های مختلف فضا را کمتر تکه‌تکه کند. نتایج:

	precision	recall	f1-score	support
0	0.73	0.67	0.70	422
1	0.56	0.35	0.43	334
2	0.65	0.70	0.67	620
3	0.64	0.57	0.60	373
4	0.54	0.71	0.62	973
5	0.69	0.51	0.58	289
6	0.72	0.67	0.70	525
7	0.58	0.51	0.54	364
8	0.76	0.83	0.79	964
9	0.63	0.50	0.56	372
accuracy			0.65	5236
macro avg	0.65	0.60	0.62	5236
weighted avg	0.65	0.65	0.65	5236

شکل ۱۳: دقت SVM بعد از tuning

	precision	recall	f1-score	support
0	0.59	0.45	0.51	422
1	0.31	0.15	0.20	334
2	0.32	0.47	0.38	620
3	0.26	0.21	0.23	373
4	0.35	0.41	0.38	973
5	0.42	0.11	0.17	289
6	0.25	0.47	0.32	525
7	0.26	0.16	0.20	364
8	0.53	0.52	0.52	964
9	0.25	0.10	0.15	372
accuracy			0.36	5236

شکل ۱۴: دقت DT بعد از tuning

فاز چهارم - Evaluation

در این فاز، مدل‌های بهینه‌شده SVM و درخت تصمیم که در فاز قبل آموزش داده و تیونینگ شده‌اند، بر روی مجموعه تست (۲۰% از دیتاست) ارزیابی شدند.

معیارهای ارزیابی

- دقت (**Accuracy**):
نسبت تعداد پیش‌بینی‌های صحیح مدل به کل نمونه‌های موجود در مجموعه‌ی تست. دقت بالا نشان‌دهنده عملکرد کلی خوب مدل است، اما در دیتاست‌های نامتوازن ممکن است گمراه‌کننده باشد، زیرا مدل می‌تواند صرفاً با پیش‌بینی کلاس غالب، دقت بالایی نشان دهد.

- **Precision**:
نسبت نمونه‌های مثبت که به درستی شناسایی شده‌اند به کل نمونه‌هایی که مدل به عنوان مثبت پیش‌بینی کرده است. دقت بالا نشان می‌دهد که مدل خطاهای مثبت کاذب (false positives) کمی دارد و پیش‌بینی‌های مثبت آن قابل اعتماد هستند. این معیار در مواقعی که هزینه خطای مثبت بالا است (مثلاً تشخیص بیماری نادر) اهمیت ویژه‌ای دارد.

- **Recall**:
نسبت نمونه‌های واقعی مثبت که توسط مدل به درستی شناسایی شده‌اند به کل نمونه‌های واقعی مثبت. Recall بالا نشان می‌دهد که مدل تعداد کمی از نمونه‌های مثبت واقعی را از دست می‌دهد. این معیار برای کلاس‌های مهم یا کم‌تعداد، مانند elephant در دیتاست ما، حیاتی است.

- **F1 Score**:
میانگین هارمونیک Precision و Recall است و تعادلی بین توانایی مدل در شناسایی نمونه‌های مثبت و دقت پیش‌بینی‌های مثبت ایجاد می‌کند. Macro-F1 میانگینی ساده (بدون وزن‌دهی) بر روی تمام کلاس‌ها است و برای دیتاست‌های نامتوازن مناسب‌تر است، زیرا به کلاس‌های کوچک نیز اهمیت برابر می‌دهد.

نتایج ارزیابی مدل‌ها

مدل‌ها از فایل‌های ذخیره‌شده فاز قبل لود شدند و پیش‌بینی‌ها بر روی تست ست انجام گرفت. نتایج به صورت ماتریس درهم‌ریختگی و معیارهای عددی استخراج شد.

تحلیل خطاهای مدل‌ها

مدل SVM

مدل SVM اشتباهات نسبتاً کمتر و متمرکزتری نسبت به درخت تصمیم دارد. مهم‌ترین موارد خطا شامل موارد زیر است:

- کلاس **cat**: بیشترین اشتباه به کلاس‌های dog (۱۲۵ مورد) و spider (۴۶ مورد). احتمالاً به دلیل شباهت ظاهری یا الگوهای مو.

- کلاس **horse**: اشتباه به dog (۸۹) و cow (۲۸) که می‌تواند ناشی از ساختار بدنی چهارپای مشابه باشد.

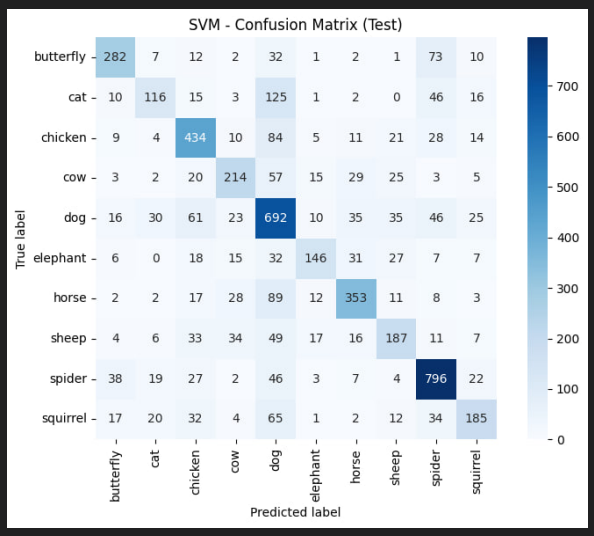
- کلاس **squirrel**: اشتباه به dog (۶۵) و spider (۳۴) احتمالاً به دلیل اندازه کوچک و ویژگی‌های ظریف که استخراج آن‌ها دشوار است.

- الگوی کلی: کلاس dog نقش یک "sink" را دارد؛ یعنی بسیاری از کلاس‌ها به آن اشتباه می‌شوند. این وضعیت می‌تواند ناشی از تعداد بیشتر تصاویر dog یا عمومی‌تر بودن ویژگی‌های ظاهری آن باشد.

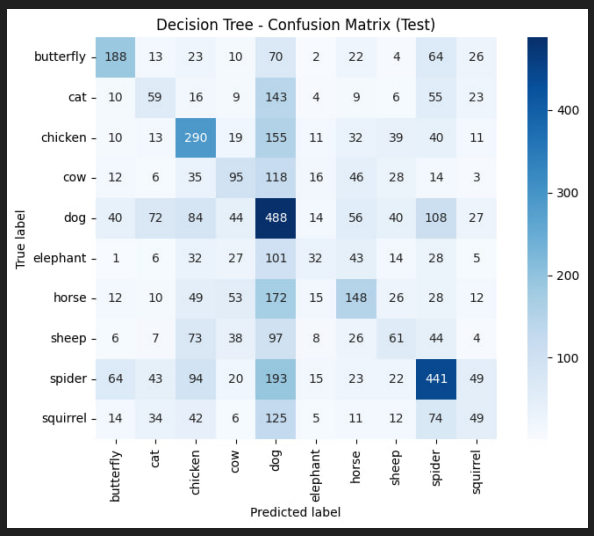
مدل Tree Decision

مدل درخت تصمیم نسبت به SVM اشتباهات بیشتر و پراکنده‌تری ایجاد کرده است. موارد زیر قابل توجه‌اند:

- کلاس **spider**: اشتباه بسیار زیاد به dog (۱۹۳ مورد) و chicken (۹۴). این ضعف نشان‌دهنده دشواری در جداسازی الگوهای بافتی عنکبوت در مدل درخت تصمیم است.
- الگوی کلی: همانند SVM، کلاس dog بیشترین مقدار اشتباهات ورودی را جذب می‌کند. اما نسبت به SVM، درخت تصمیم در کلاس‌هایی مثل spider و horse خطاهای بسیار بیشتری دارد.



شکل ۱۵: ماتریس درهم‌ریختگی مدل SVM



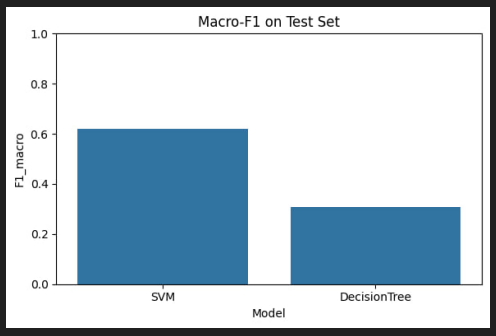
شکل ۱۶: ماتریس درهم‌ریختگی مدل DT

کلاس	SVM Recall	SVM Precision	SVM F1	DT Recall	DT Precision	DT F1
butterfly	۶۶.۸۲٪	۷۲.۸۷٪	۶۹.۷۲٪	۴۴.۵۵٪	۵۲.۶۶٪	۴۸.۲۷٪
cat	۳۴.۷۳٪	۵۶.۳۱٪	۴۲.۹۶٪	۱۷.۶۶٪	۲۲.۴۳٪	۱۹.۷۷٪
chicken	۷۰.۰۰٪	۶۴.۸۷٪	۶۷.۳۴٪	۴۶.۷۷٪	۳۹.۳۰٪	۴۲.۷۱٪
cow	۵۷.۳۷٪	۶۳.۸۸٪	۶۰.۴۵٪	۲۵.۴۷٪	۲۹.۶۰٪	۲۷.۳۸٪
dog	۷۱.۱۲٪	۵۴.۴۵٪	۶۱.۶۸٪	۵۰.۱۵٪	۲۹.۳۶٪	۳۷.۰۴٪
elephant	۵۰.۵۲٪	۶۹.۱۹٪	۵۸.۴۰٪	۱۱.۰۷٪	۲۶.۲۳٪	۱۵.۵۷٪
horse	۶۷.۲۴٪	۷۲.۳۴٪	۶۹.۶۹٪	۲۸.۱۹٪	۳۵.۵۸٪	۳۱.۴۶٪
sheep	۵۱.۳۷٪	۵۷.۸۹٪	۵۴.۴۴٪	۱۶.۷۶٪	۲۴.۲۱٪	۱۹.۸۱٪
spider	۸۲.۵۷٪	۷۵.۶۷٪	۷۸.۹۷٪	۴۵.۷۵٪	۴۹.۲۲٪	۴۷.۴۲٪
squirrel	۴۹.۷۳٪	۶۲.۹۳٪	۵۵.۵۶٪	۱۳.۱۷٪	۲۳.۴۴٪	۱۶.۸۷٪

شکل ۱۷: نتیجه بدست آماده از ماتریس درهم‌ریختگی

	model	split	accuracy	precision_macro	recall_macro	f1_macro
0	SVM	train	0.998520	0.998687	0.998234	0.998458
1	SVM	test	0.650306	0.650393	0.601483	0.619203
2	DecisionTree	train	0.516020	0.521037	0.448911	0.468382
3	DecisionTree	test	0.353514	0.332024	0.299552	0.306279

شکل ۱۸: نتیجه ارزیابی ها



شکل ۱۹: مقایسه Macro-F1 مدل‌ها بر روی تست ست

نمودار Macro F1-Score روی داده‌های تست، عملکرد دو مدل SVM و Decision Tree را در طبقه‌بندی ده کلاس مقایسه می‌کند. مدل SVM با امتیاز تقریبی ۶.۰ (معادل ۶۰%) عملکرد قابل توجهی بهتر از Decision Tree با مقدار حدود ۳.۰ (معادل ۳۰%) داشته است. این اختلاف نشان می‌دهد که SVM توانایی بیشتری در ایجاد تعادل میان معیارهای دقت و بازخوانی برای کلاس‌های مختلف دارد؛ به‌طور خاص، کلاس‌هایی مانند spider و chicken دارای Recall بالای ۷۰% هستند.

در مقابل، مدل Decision Tree در کلاس‌های پیچیده‌ای مانند elephant و squirrel عملکرد ضعیفی نشان داده (Recall کمتر از ۱۵%) و بیش‌ازحد مستعد overfitting بوده است. این رفتار منجر به اشتباهات گسترده‌ای در پیش‌بینی شده که نمونه بارز آن اشتباه طبقه‌بندی ۱۹۳ تصویر spider به عنوان dog است.

در مجموع، این نتایج بیانگر آن است که SVM انتخاب مناسب‌تری برای داده‌های تصویری با شباهت‌های بصری میان کلاس‌ها محسوب می‌شود؛ در حالی که برای بهبود عملکرد Decision Tree لازم است مثلا Tuning یا مدل‌های Ensemble (نظیر Random Forest و Gradient Boosting) استفاده شود تا قابلیت تعمیم آن افزایش یابد.

فاز پنجم - Image Segmentation, Unimportant Region Removal

ایده اصلی به این صورته که هر ناحیه از تصویر به طور جداگانه حذف یا mask می‌شود و میزان کاهش اعتماد مدل به کلاس درست اندازه‌گیری می‌گردد. نواحی‌ای که بیشترین کاهش اعتماد را ایجاد کنند، به عنوان بخش‌های مهم و تأثیرگذار تصویر شناخته می‌شوند.

این روش باعث:

- برجسته‌سازی نواحی معنادار و مرتبط با طبقه‌بندی،
- حذف بخش‌های غیرضروری و کاهش نویز،
- ساده‌تر شدن تصویر برای فازهای بعدی مانند fine-tuning،

می‌شود.

مزیت اصلی این رویکرد آن است که اهمیت نواحی بر اساس رفتار واقعی مدل تعیین می‌شود، نه براساس ویژگی‌های ثابت پیکسل. البته محاسبه حساسیت برای تمام نواحی سنگین است و برای تصاویر بزرگ نیاز به بهینه‌سازی‌هایی مانند پردازش موازی دارد.

مرحله Segmentation تولید سوپرپیکسل‌ها با الگوریتم SLIC

در این مرحله، تصویر توسط الگوریتم SLIC به مجموعه‌ای از سوپرپیکسل‌های کوچک و همگن تقسیم می‌شود. این الگوریتم با خوشه‌بندی در فضای رنگی Lab و مختصات مکانی، نواحی‌ای تولید می‌کند که معمولاً لبه‌های طبیعی شیء را بهتر حفظ می‌کنند.

تقریباً ۳۰۰–۲۵۰ سوپرپیکسل برای هر تصویر ایجاد می‌شود تا تعادلی میان دقت ناحیه‌بندی و هزینه محاسباتی برقرار شود. از مزایای این روش، حفظ بافت و ساختار محلی تصویر است؛ اما پارامتر compactness باید با دقت تنظیم شود، زیرا مقدار زیاد باعث نواحی بیش‌ازحد منظم و مقدار کم موجب نواحی نامنظم و نویزی می‌شود.

هر سوپرپیکسل با یک بردار ۶بعدی شامل میانگین رنگ Lab، موقعیت نرمال‌شده مرکز و نسبت اندازه توصیف می‌شود. این ویژگی‌ها شباهت‌های بصری، روابط فضایی و اهمیت نواحی بزرگ‌تر را کدگذاری می‌کنند و خوشه‌بندی را ساده و مؤثر می‌سازند. تحلیل نشان می‌دهد که این بردارها تعادل خوبی بین تفسیرپذیری و کارایی محاسباتی ایجاد می‌کنند، اگرچه در تصاویر با گرادیان‌های رنگی پیچیده ممکن است جزئیات ظریف کمی از دست برود.

خوشه‌بندی دو مرحله‌ای

برای ادغام سوپرپیکسل‌ها به نواحی بزرگ‌تر (حدود ۱۰–۲۰ ناحیه نهایی)، یک خوشه‌بندی هیبریدی دو مرحله‌ای طراحی شد: ابتدا DBSCAN برای شناسایی خوشه‌های چگال در فضای ویژگی (پس از نرمال‌سازی استاندارد)، و سپس Agglomerative Clustering روی نماینده‌های (میانگین) خوشه‌های DBSCAN.

DBSCAN، به عنوان مرحله اول، نواحی پرچگالی (مانند بدنه حیوان) را بدون نیاز به پیش‌تعریف تعداد خوشه‌ها جدا می‌کند و نقاط پراکنده (نویز پس‌زمینه) را حذف می‌کند این ویژگی، اون را برای دیتاست‌های نویزی مثل تصاویر حیوانات در محیط‌های طبیعی ایده‌آل میکند، زیرا پس‌زمینه‌های نامنظم (مثل برگ‌ها) را به عنوان نویز طبقه‌بندی می‌کند.

ولی گاها در بعضی موارد مثل شکل ۲۱ هم مدل خوب تصمیم نمیگیره و پس‌زمینه رو به عنوان مورد مهم نگه میداره و حیوان رو حذف میکنه.

مرحله دوم (Agglomerative با linkage Ward)، نماینده‌ها را به تعداد ثابت ادغام می‌کند، که واریانس داخل‌نواحی را مینیمم می‌نماید و نواحی همگن بزرگ‌تری (مانند "سر" یا "پاها") تولید می‌کند.

این ترکیب، مزایای DBSCAN (انعطاف‌پذیری) و Agglomerative (کنترل اندازه خروجی) را ادغام می‌کند، اما چالش آن در تنظیم eps است: مقادیر کوچک خوشه‌های بیش از حد ریز ایجاد می‌کند، در حالی که مقادیر بزرگ، تمایز شیء از پس‌زمینه را کاهش می‌دهد تحلیل تجربی نشان می‌دهد که این روش، نسبت به k-means تک‌مرحله‌ای، مرزهای معنادارتر و کمتر over-segmented تولید می‌کند. همچنین نیاز به تنظیم کردن k در هر تصویر ندارد.

در تصویر ۲۲، چون SLIC با compactness بالا سوپرپیکسل‌های درشت و همگن تولید می‌کند که منجر به K_{FINAL} پایین (حدود ۱۰ ناحیه) می‌شود و تمایز کمی بین شیء و پس‌زمینه ایجاد می‌کند یعنی نواحی حاشیه‌ای (مثل لبه‌های علف) به عنوان بخشی از خوشه‌های چگال DBSCAN گروه‌بندی می‌شوند و جدا نمی‌مانند. علاوه بر این، Agglomerative با linkage Ward ادغام محافظه‌کارانه‌ای انجام می‌دهد و نواحی کوچک را حفظ می‌کند، در حالی که دلتای SVM برای این نواحی ناچیز است (حذفشان افت کمی در اعتماد مدل ایجاد می‌کند، چون تصویر واضح و bias مدل به ویژگی‌های کلی مثل رنگ قهوه‌ای-سبز زیاد است). نتیجه ماسک‌ها محدود به لکه‌های کوچک می‌مانند و حذف گسترده‌ای رخ نمی‌دهد برای بهبودش، DBSCAN eps رو باید کوچک‌تر یا n_{keep} کمتر بشه تا pruning تهاجمی‌تری داشته میبودیم. اما خطر از دست دادن جزئیات کلیدی (مثل پاها) رو دار.

شکل ۲۰: نمونه کلاسترینگ خوب- تصویر اول

شکل ۲۱: نمونه کلاسترینگ خوب- تصویر دوم

شکل ۲۲: نمونه کلاسترینگ بد- تصویر اول

شکل ۲۳: نمونه کلاسترینگ بد- تصویر دوم

محاسبه اهمیت نواحی با مدل SVM

اهمیت هر ناحیه بر اساس "دلتا امتیاز" تعریف می‌شود: امتیاز اعتماد SVM به کلاس غالب تصویر (S_{orig}) منهای امتیاز پس از ماسک کردن آن ناحیه (S^k_{mask}). ماسک کردن با جایگزینی محتوای ناحیه با رنگ ثابت (خاکستری)، اطلاعات بصری را بدون تغییر ساختار هندسی حذف می‌کند . این شبیه‌سازی "حذف آزمایشی"، حساسیت مدل را مستقیماً اندازه‌گیری می‌کند.

اگر دلتا مثبت و بزرگ باشد، ناحیه کلیدی است (حذف آن مدل را گیج می‌کند)، که منطقاً نواحی discriminative (مانند چشم فیل یا بال پروانه) را اولویت می‌دهد. این رویکرد، نسبت به روش‌های saliency map های مبتنی بر گرادیان (مانند Grad-CAM)، تفسیرپذیرتر است زیرا مستقیماً به خروجی

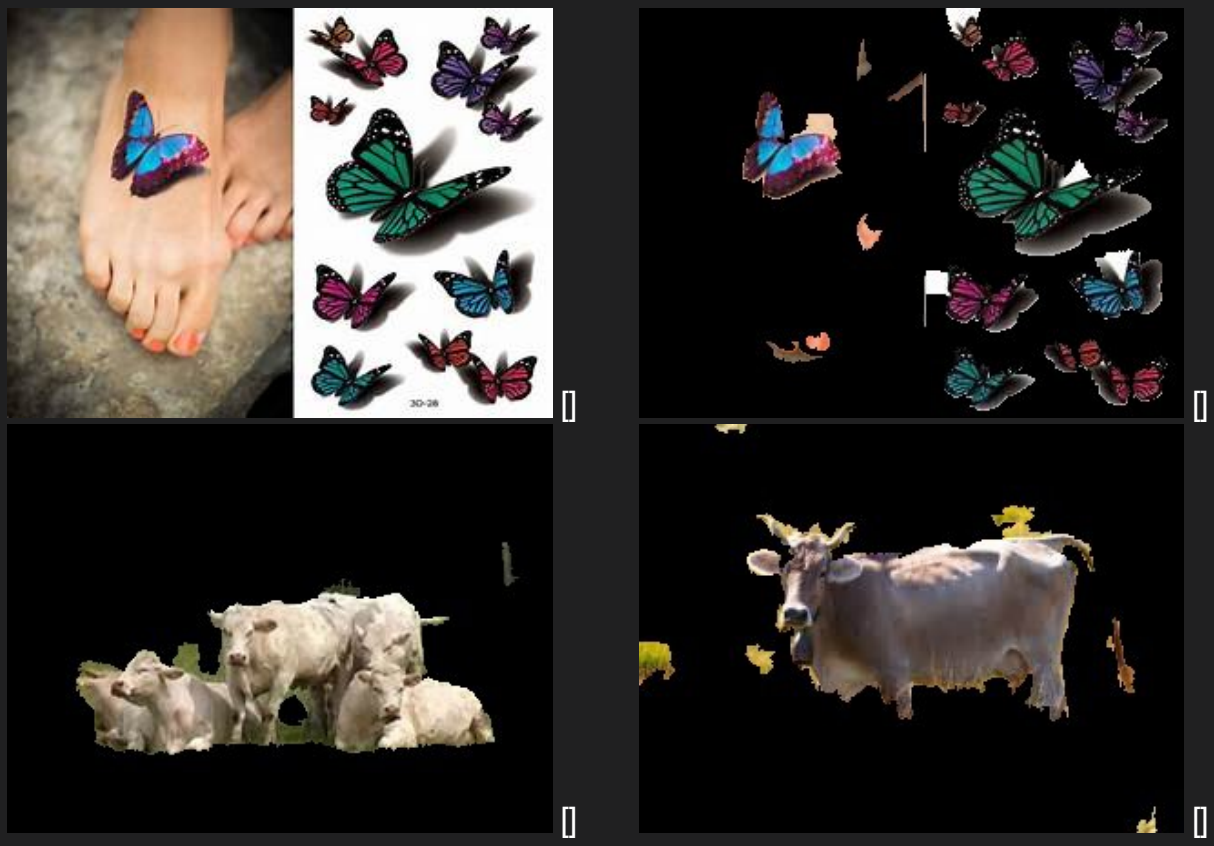
مدل وابسته است و کمتر به معماری شبکه‌های عمیق نیاز دارد؛ علاوه بر این، در دیتاست‌های کوچک مانند این پروژه، از overfit جلوگیری می‌کند. با این حال، وابستگی به SVM ممکن است bias مدل (مانند تمایل به کلاس‌های غالب مانند سگ) را منعکس کند، و محاسبات تکراری (برای هر ناحیه) زمان‌بر است

این روش، توضیح‌پذیری را افزایش می‌دهد و نواحی را بر اساس "تأثیر واقعی" انتخاب می‌کند، نه فرضیات بصری.

انتخاب و ماسک نواحی غیرمهم

در نهایت، n_{keep} ناحیه با بیشترین دلتا (معمولاً ۳–۴) حفظ می‌شود، و بقیه ماسک می‌گردند تا تصویر ساده‌شده‌ای تولید شود. این تصمیم‌گیری، مفهوم "تمرکز روی essentials" را عملی می‌سازد: حفظ نواحی کلیدی، مدل را از distractions (پس‌زمینه یا جزئیات غیرضروری) رها می‌کند و generalization را بهبود می‌بخشد، به ویژه در فاز ۶ که این تصاویر ماسک‌شده به عنوان ورودی جدید استفاده می‌شوند.

مزیت آن، کاهش حجم داده بدون از دست دادن اطلاعات discriminative است، که آموزش سریع‌تر و دقت بالاتر را تضمین می‌کند؛ اما خطر، over-pruning است – اگر n_{keep} خیلی کوچک باشد، جزئیات حیاتی (مانند اندام‌های کوچک عنکبوت) حذف می‌شود. تحلیل نشان می‌دهد که این مرحله، مؤثر بین segmentation و fine-tuning است، و با تنظیم پویا n_{keep} بر اساس پیچیدگی تصویر، تعادل بهینه‌ای ایجاد می‌کند.



شکل ۲۴: نمونه هایی از کلاسترینگ

فاز ششم - Train2

مراحل مشابه فاز۳ هم در این فاز انجام شد بااین تفاوت که برای داده Train از عکس های پردازش شده در فاز قبلی استفاده شد.

```
accuracy                    0.60
```

شکل ۲۵: دقت SVM

```
accuracy                    0.29
```

شکل ۲۶: دقت DT

```
Best SVM parameters: {'C': np.float64(84.31013932082456), 'gamma': np.float64(0.0006290644294586153), 'kernel': 'rbf'}
Best DT parameters: {'criterion': 'gini', 'max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 15}
```

شکل ۲۷: پارامترهای بهینه‌سازی شده مدل‌های SVM و Decision Tree

فاز هفتم - Evaluation2

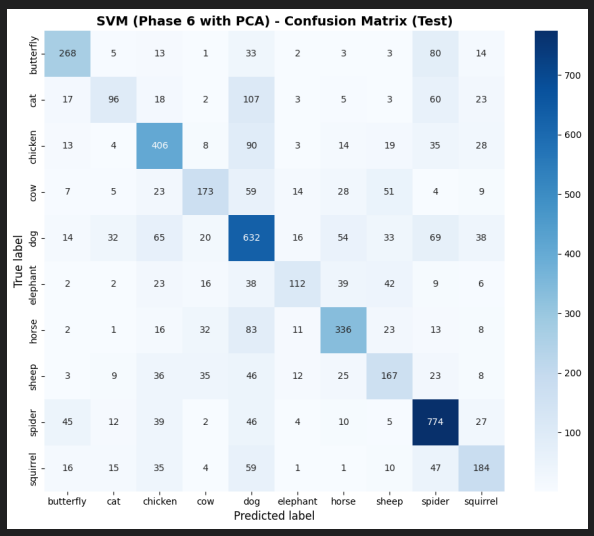
نتایج ارزیابی مدل‌ها

	precision	recall	f1-score	support
butterfly	0.6925	0.6351	0.6625	422
cat	0.5304	0.2874	0.3728	334
chicken	0.6024	0.6548	0.6275	620
cow	0.5904	0.4638	0.5195	373
dog	0.5298	0.6495	0.5836	973
elephant	0.6292	0.3875	0.4797	289
horse	0.6524	0.6400	0.6462	525
sheep	0.4691	0.4588	0.4639	364
spider	0.6948	0.8029	0.7449	964
squirrel	0.5333	0.4946	0.5132	372
accuracy			0.6012	5236
macro avg	0.5924	0.5475	0.5614	5236
weighted avg	0.6000	0.6012	0.5941	5236

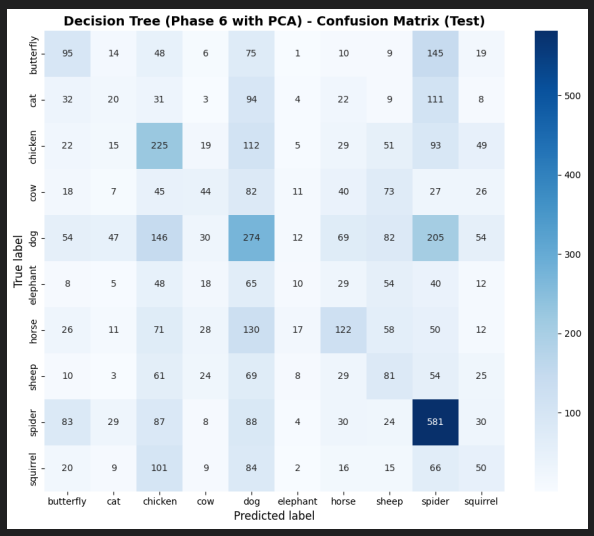
شکل ۲۸: دقت SVM بعد از tuning

	precision	recall	f1-score	support
butterfly	0.2582	0.2251	0.2405	422
cat	0.1250	0.0599	0.0810	334
chicken	0.2607	0.3629	0.3034	620
cow	0.2328	0.1180	0.1566	373
dog	0.2554	0.2816	0.2678	973
elephant	0.1351	0.0346	0.0551	289
horse	0.3081	0.2324	0.2649	525
sheep	0.1776	0.2225	0.1976	364
spider	0.4235	0.6027	0.4974	964
squirrel	0.1754	0.1344	0.1522	372
accuracy			0.2869	5236
macro avg	0.2352	0.2274	0.2217	5236
weighted avg	0.2648	0.2869	0.2671	5236

شکل ۲۹: دقت DT بعد از tuning



شکل ۳۰: ماتریس درهم‌ریختگی مدل SVM



شکل ۳۱: ماتریس درهم‌ریختگی مدل DT

تحلیل افت دقت پس از حذف خوشه‌ها

وقتی دقت مدل از حدود ۶۵% به ۶۰% کاهش پیدا کرد، فکر میکنیم به این دلایل افت کرده:

- معیار "کم‌اهمیت" همیشه دقیق نیست: ما در فاز ۵ خوشه‌هایی رو حذف کردیم که به نظر مدل فاز ۳ کم‌ترین تأثیر رو روی پیش‌بینی‌ها داشتن. اما خود مدل دقتش حدود ۶۵% هست و بعضی پیش‌بینی‌ها اشتباه یا نویزی هستن. بنابراین ممکنه خوشه‌هایی که واقعاً مهم بودن هم حذف شدن، فقط چون مدل اون‌ها رو کم‌اثر می‌دید.

- برخی نواحی حتی اگر به نظر اضافه باشن، اطلاعات مفیدی دارن: مثلاً پس‌زمینه‌ی تصاویر حیوانات (چمن، شاخه‌ها، تار عنکبوت و ...) گاهی به تفکیک کلاس‌ها کمک می‌کنه. وقتی این نواحی حذف شدن، بخشی از سیگنال‌های تمایزبخش از بین رفت و مدل روی بعضی نمونه‌ها دچار اشتباه شد.

- ماسک یا کراپ روی فیچرها اثر منفی گذاشت: فیچرهایی که استفاده کردیم مثل آمار رنگ، HSV، GLCM، لبه‌ها، LBP، هیستوگرام رنگ و HOG هستن. وقتی بخشی از تصویر ماسک یا کراپ می‌شه، مرزهای مصنوعی یا بخش‌های یک‌رنگ ایجاد می‌شه که بعضی فیچرها مثل HOG و لبه‌ها رو تحت تأثیر قرار می‌ده. علاوه بر این، resize بعد از کراپ باعث می‌شه ساختار گرادیان‌ها تغییر کنه و HOG آسیب ببینه.

فاز هشتم - نمره اضافه

برای حذف بخش‌های نامربوط در تصاویر تست بدون برچسب، ابتدا تصویر را با SLIC به سوپرپیکسل‌ها تقسیم کردیم و برای هر سوپرپیکسل ویژگی‌های ساده‌ای مثل میانگین رنگ در فضای LAB و مختصات مرکز نرمال‌شده استخراج شد. سپس سوپرپیکسل‌های مشابه با DBSCAN خوشه‌بندی شدند و از هر خوشه یک نماینده انتخاب کردیم. روی این نماینده‌ها خوشه‌بندی سلسله‌مراتبی (Ward) اعمال شد تا نواحی نهایی شکل بگیرند. برای تعیین اهمیت هر ناحیه، هر ناحیه را به‌نوبت ماسک کردیم و کاهش احتمال پیش‌بینی مدل مرجع (DT/SVM) روی تصویر را اندازه‌گیری کردیم. نواحی با بیشترین کاهش احتمال به عنوان مهم نگه داشته شدند و سایر نواحی ماسک شدند تا بخش‌های نامربوط حذف شوند. در نهایت، نسخه‌های ماسک‌شدهٔ تصاویر تست را با همان خط لولهٔ فاز ۷ ارزیابی کردیم.