

به نام خدا

گزارش فاز دوم پروژه هوش محاسباتی

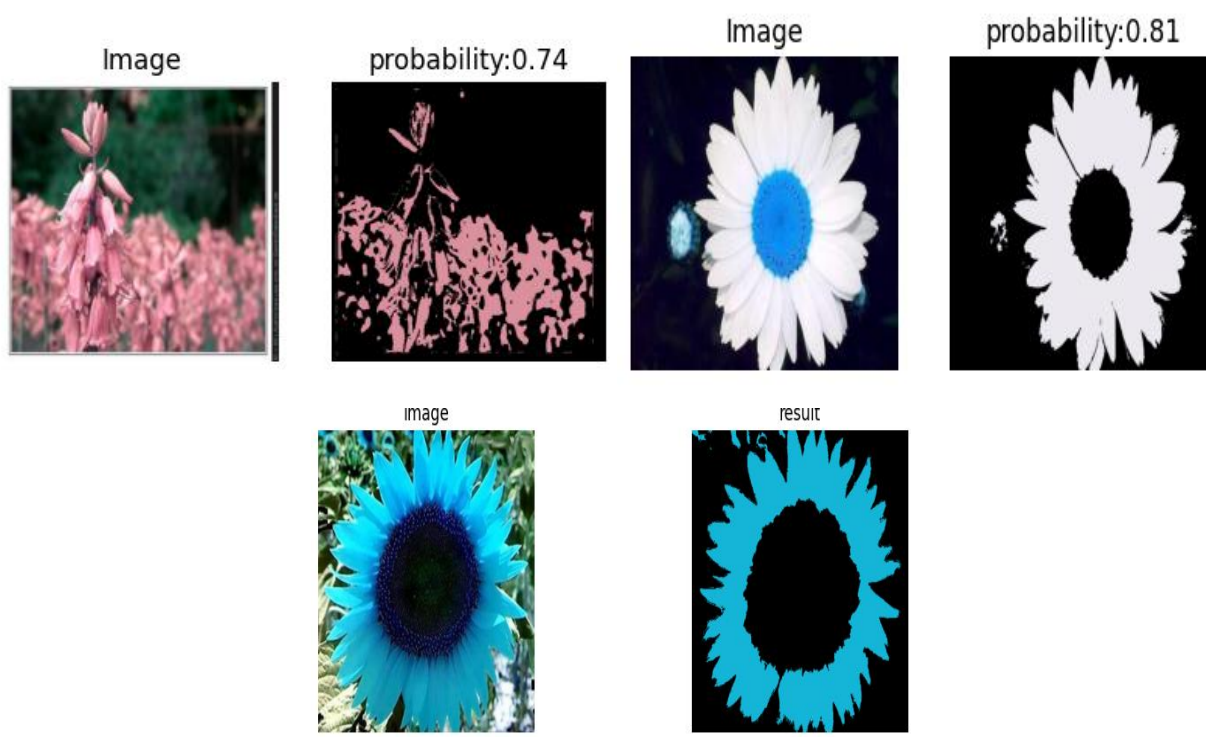
تینا توکلی - هادی امینی

بخش اول:

در بخش اول پروژه، قصد داریم نواحی ای را پیدا کنیم که بیشترین اطلاعات را در خصوص طبقه بندی تصویر به ما می‌دهند. برای انجام این بخش اینگونه عمل میکنیم:

در ابتدا برای هر عکس بردار میانگین رو میگیریم و سپس دقت را با توجه به میانگین بردارها و لیبل های عکس با استفاده از RandomForrestClassifier به دست می‌آوریم .
یعنی در ابتدا برای هر عکس بردار میانگین 4 ناحیه به دست آمده را حساب میکنیم سپس مدل را با کل بردارها train میکنیم ، به ازای هر عکس یکی از نواحی رو حذف کرده و درصد احتمال تعلق به کلاس درست با توجه به لیبل محاسبه میشود و سپس ناحیه ای که با حذف ان به درصد بالاتری میرسیم را به عنوان کاندید حذف در نظر میگیریم و این کار را برای تمامی عکسها تکرار میکنیم و سپس بعد از بررسی تمام عکسها نواحی انتخاب شده را حذف میکنیم و بردار میانگین جدید را حساب کرده و مجدد classifier را آموزش میدهیم ،این کار را تا زمانی که به یک ناحیه برسیم تکرار میکنیم.

چند نمونه از حذف درست نواحی(بهترین تشخیص مدل):



تصویر 1 - نمونه هایی از تشخیص ناحیه اصلی

چند نمونه از حذف نادرست نواحی (بیشترین اشتباهات مدل):



تصویر 2 - نمونه هایی از تشخیص نادرست

روند اجرای کد بخش اول :

```
i=553, k=2
Accuracy: 26.00%
Accuracy: 31.00%
i=554, k=2
Accuracy: 18.00%
Accuracy: 34.00%
i=555, k=2
Accuracy: 60.00%
Accuracy: 12.00%
i=556, k=2
Accuracy: 21.00%
Accuracy: 18.00%
i=557, k=2
Accuracy: 39.00%
Accuracy: 16.00%
i=558, k=2
Accuracy: 12.00%
Accuracy: 21.00%
i=559, k=2
Accuracy: 26.00%
Accuracy: 29.00%
----final----
Accuracy: 74.11%
```

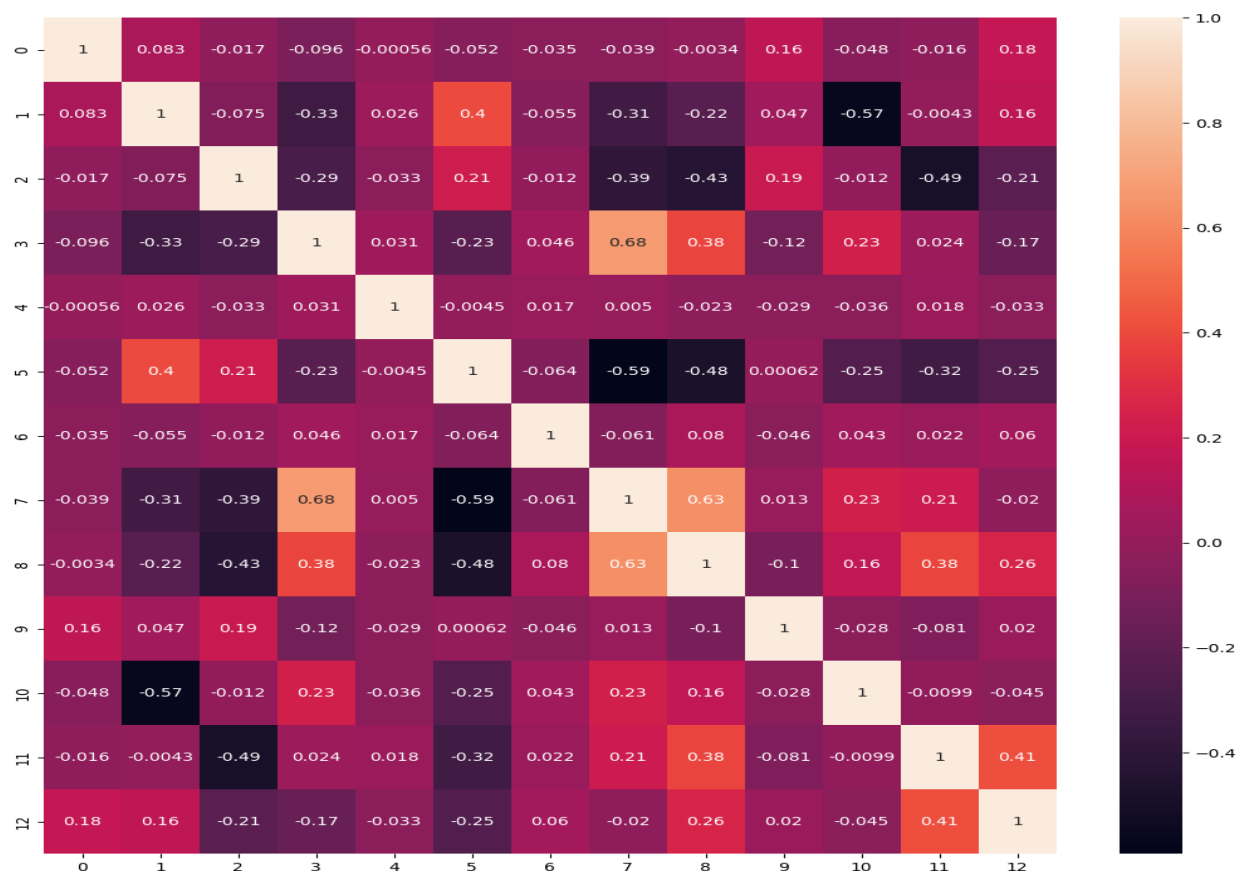
تصویر 3 - محاسبه دقت پس از حذف نواحی

توضیحات مربوط به ویژگیها:

- (1) **واریانس h, s, v :** نیاز به محاسبه پراکندگی رنگ برای تشخیص نواحی مشابه البته ویژگی مناسبی نیست زیرا ما براساس رنگ گل هم دسته بندی میکنیم و همچنین طیف و شدت آن پس توجه به پراکندگی تنها باعث بدتر دسته بندی شدن گلها میشود.
- (2) **میانگین h و مینیمم و ماکسیمم طیف h :** اصلی ترین شاخصه جداسازی گلها رنگ آنهاست و معیار h رنگ را نشان میدهد.
تاثیر این 3 بر دسته بندی درست گلها: بیشترین تاثیر را h و سپس s و در آخر v است.
البته ممکن است استفاده از مینیمم و ماکسیمم باعث شود گاهی نویزها کار را خراب میکنند.
- (3) **میانگین فاصله از مرکز و زاویه قرارگیری در عکس:** نیاز به موقعیت مکانی داریم برای تشخیص گلها از هم، همچنین به دلیل اینکه عموماً گلهای داخل یک دسته از لحاظ سایز مشابه هستند بهترین حالت در صورتی به وجود میاد که ضریب یکسان با میانگین h داشته باشد (زیرا با یک نسبت باعث جداسازی و لیبل گذاری داده ها میشوند).
- (4) **شباهت به دایره:** هدف اصلی ما دسته بندی گلهاست بنابراین اهمیت شکل گل ها نیز برای ما مهم تر است که این معیار، معیار درست تری نسبت به مقایسه محیط و مساحت است و یا نسبت به میانگین زاویه قرارگیری گلها.
- (5) **aspect-ratio:** کاری که انجام میدهیم نسبت طول به عرض بود که با توجه به پراکندگی قرارگیری گلها در عکس، تقریباً مختصات خود عکس را برمیگرداند بنابراین ویژگی کاملی نیست.
- (6) **مساحت:** ممکن است در چند عکس یک گل باشد ولی با تعداد و یا سایز متفاوت که ممکن است در عکس های مختلف، متفاوت باشد البته معیار کاملی نیست.
- (7) **مرکز کلاستر:** احتمال دارد دو کلاستر با پراکندگی نقاط زیاد و کم مرکز یکسانی داشته باشند در حالی که یک دسته نیستند. (به علت پراکندگی زیاد به احتمال بیشتری گلهای یکسان را در دسته های متفاوتی قرار میدهد)

بخش دوم:

بخش دوم پروژه، به انتخاب ویژگی اختصاص دارد بنابراین پس از بدست آوردن رتبه بندی ناحیه های مختلف تصویر، به سراغ مهم ترین ناحیه هر تصویر میرویم .
در هر مرحله یک ویژگی را انتخاب نموده و دوباره مدل را با 80 درصد داده آموزش می دهیم ، دقت را می گیریم و کرولیشن آن را نیز حساب میکنیم (کرولیشن برابر است با ماکسیم کرولیشن ها با ویژگی هایی که تا به الان انتخاب شده است)



تصویر 4- کرولیشن بین 13 ویژگی

وسپس با توجه به فرمول زیر امتیاز آن را حساب میکنیم:

$$score = \frac{2}{\frac{1}{acc} + \frac{1}{1-correlation}}$$

این کار را برای سایر ویژگیها نیز انجام می دهیم و در اخر از بین تمام این ها، ویژگی که بیشترین امتیاز را دارا باشد انتخاب میشود و این کار را تکرار میکنیم تا زمانی که به شرط توقف برسیم.

روند اجرای کد بخش دوم :

```
selected count: 6 - checking feature 1
Accuracy: 68.81%
selected count: 6 - checking feature 3
Accuracy: 64.22%
selected count: 6 - checking feature 6
Accuracy: 66.97%
selected count: 6 - checking feature 7
Accuracy: 64.22%
selected count: 6 - checking feature 9
Accuracy: 64.22%
selected count: 6 - checking feature 10
Accuracy: 73.39%
selected count: 6 - checking feature 12
Accuracy: 65.14%
feature 10 selected

selected count: 7 - checking feature 1
Accuracy: 66.97%
selected count: 7 - checking feature 3
Accuracy: 66.97%
selected count: 7 - checking feature 6
Accuracy: 69.72%
selected count: 7 - checking feature 7
Accuracy: 66.06%
selected count: 7 - checking feature 9
Accuracy: 67.89%
selected count: 7 - checking feature 12
Accuracy: 66.06%
selected_feats [0, 8, 11, 2, 4, 5, 10]
acc 73.39%
```

تصویر 5- انتخاب ویژگی و محاسبه دقت

ویژگیها:

```
['h_average', 's_average', 'v_average', 'distance_average', 'angle_average', 'circularity', 'aspect_ratio', 'perimeter', 'distance_variance', 'angle_variance', 'h_variance', 's_variance', 'v_variance']
```

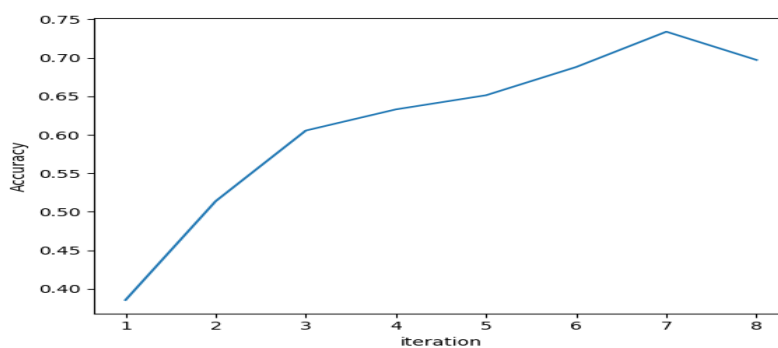
نتیجه نهایی انتخاب ویژگیها :

```
Ranking of features:
1.h_average
2.distance_variance
3.s_variance
4.v_average
5.angle_average
6.circularity
7.h_variance
```

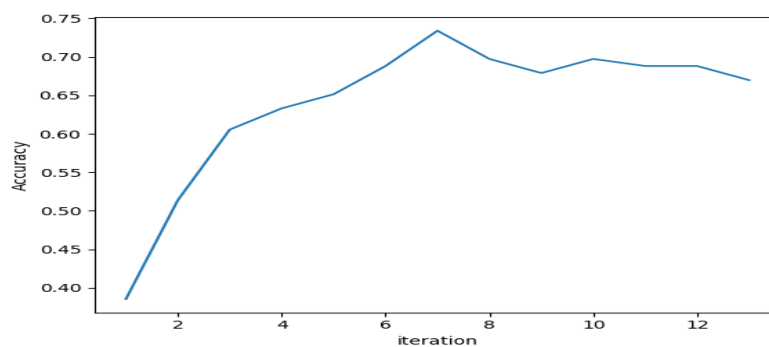
تصویر 6- ویژگی های برتر

توضیحات شرط توقف:

در مقایسه دقت هر مرحله (انتخاب ویژگی) شرط را برای زمانی گذاشتیم که دقت بیش از 3 درصد افت کند یا 2 بار متوالی نزولی باشد.



تصویر 7- پلات در صورت داشتن شرط توقف



تصویر 8- پلات در صورت نداشتن شرط توقف