

ABSTRACT

L'idea di questo progetto è quella di classificare 5 tipologie di fiori (Margherita, Dente di Leone, Girasole, Rosa e Tulipano) attraverso un modello di Apprendimento Supervisionato.

DATASET

Il dataset contiene 4526 immagini di fiori precedentemente citati, raggruppati in 5 cartelle (ognuna dedicata ad ogni tipologia):

- 818 immagini per la tipologia 'Margherita';
- 1054 immagini per la tipologia 'Dente di Leone';
- 828 immagini per la tipologia 'Rosa';
- 804 immagini per la tipologia 'Girasole';
- 1022 immagini per la tipologia 'Tulipano'.

FEATURES

Per quanto riguarda le features, vengono estratte 4 features per ogni esempio, ovvero:

- **Color Histogram**, che quantifica il colore del fiore in analisi, per poi essere riportato in un istogramma;
- **Hu Moments**, che quantifica la forma del fiore in analisi;
- **Haralick Texture**, che quantifica la texture del fiore in analisi;
- **HOG Histogram**, che conta le occorrenze dell'orientamento del gradiente in porzioni localizzate di un'immagine in analisi, per poi essere descritte in un istogramma.

Inoltre, viene fatto un 'flip' orizzontale e verticale per ogni esempio, per poi calcolarne l'HOG, in modo da avere un miglioramento nel livello d'informazione della feature HOG Histogram.

MODELLO ML UTILIZZATI

Il modello di 'Apprendimento Supervisionato' utilizzato per la classificazione è il '**Random Forest**', ovvero un insieme 'bagging' di Alberi di Decisione.

La predizione del Random Forest è calcolata come modalità delle previsioni fatte dagli alberi decisionali. Notare che ogni albero della foresta è addestrato su un sottoinsieme delle caratteristiche. Il Random Forest corregge l'abitudine degli alberi decisionali di andare in 'overfitting' sul loro training set.

Scelta degli Hyperparameters

Per l'ottimizzazione degli Hyperparameters viene utilizzato un '**GridSearchCv**', ovvero una ricerca esaustiva dei parametri di un estimatore.

VALUTAZIONE

Il dataset è splittato in modo da avere un 75% per il training set, un 25% per il test set.

Per quanto riguarda la valutazione, in seguito verranno mostrati i valori delle metriche di valutazione utilizzate:

(0: Margherita / 1: Dente di Leone / 2: Rosa / 3: Girasole / 4: Tulipano

```
evaluation Classification report:
```

	precision	recall	f1-score	support
0	0.71	0.61	0.66	216
1	0.70	0.80	0.75	271
2	0.67	0.54	0.60	215
3	0.72	0.79	0.76	175
4	0.65	0.69	0.67	255
avg / total	0.69	0.69	0.69	1132

- **Precision:** 69-70 %;
- **Recall:** 69-70%;
- **F1-Measure:** 69-70%.