

1) Why split classwise instead of the whole data?

A random split of 50-25-25 might result in a heavily skewed data where some classes would be over or under represented. Hence, splitting classwise would be a better representation of the data as there is an even distribution of samples from every class, reducing bias.

2) Validation accuracies which lead to the selected C

```
C_list = [0.01, 0.1, np.power(0.1, 0.5), 1, np.power(10, 0.5), 10, np.power(100, 0.5)]
score = {}
for C in C_list:
    svc = OneVsRestClassifier(LinearSVC(random_state=1000, C = C))
    svc.fit(X_train, Y_train)
    score[svc.score(X_val, Y_val)] = C
    print('Accuracy of classifier on test set with C = %f: %f' % (C, svc.score(X_val, Y_val)))

print('best C : {}'.format(score[max(score)]))
# .format(svc.score(X_val, Y_val)))
# cv = [(slice(None), slice(None))]
# clf = GridSearchCV(svc, param_grid = {'estimator__C': [0.01]}, cv=cv)
# clf.fit(X_train, Y_train)
```

```
Accuracy of classifier on test set with C = 0.010000: 0.923529
Accuracy of classifier on test set with C = 0.100000: 0.923529
Accuracy of classifier on test set with C = 0.316228: 0.926471
Accuracy of classifier on test set with C = 1.000000: 0.923529
Accuracy of classifier on test set with C = 3.162278: 0.920588
Accuracy of classifier on test set with C = 10.000000: 0.920588
Accuracy of classifier on test set with C = 10.000000: 0.920588
best C : 0.31622776601683794
```

3) Final test accuracy

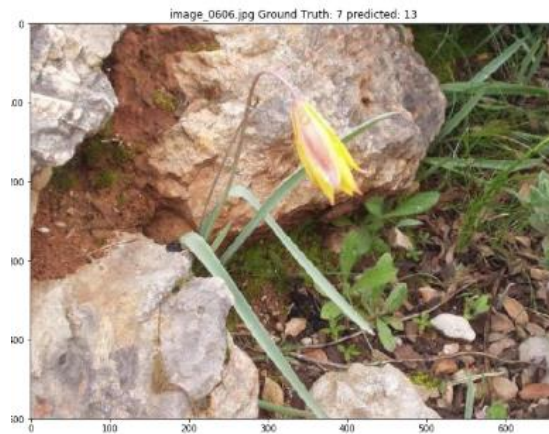
```
print('Accuracy of classifier on training set: {:.2f}'
      .format(svc.score(X_train_val, Y_train_val)))
print('Accuracy of classifier on test set: {}'.format(svc.score(X_test, Y_test)))

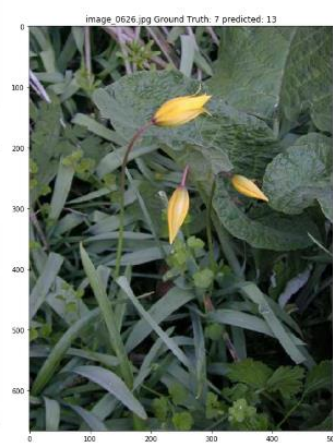
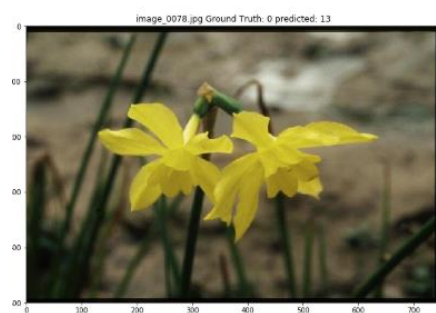
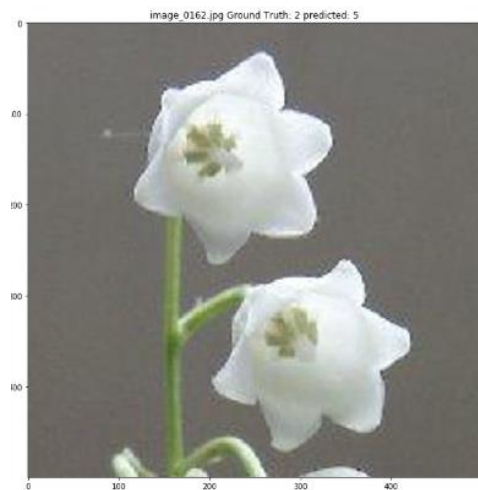
#OR
from sklearn.metrics import accuracy_score
Y_pred = svc.predict(X_test)
print('Accuracy of classifier on test set:', accuracy_score(Y_test, Y_pred))
print(Y_test)
print(Y_pred)
```

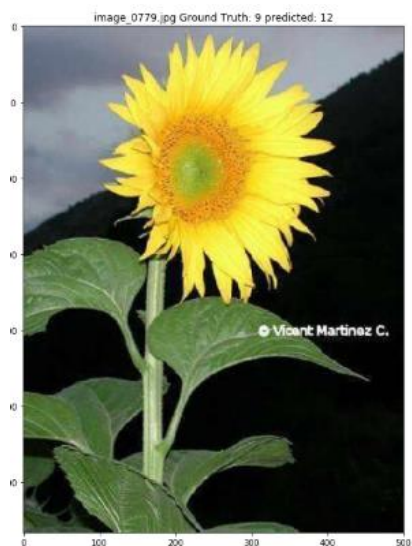
```
Accuracy of classifier on training set: 1.00
Accuracy of classifier on test set: 0.9264705882352942
Accuracy of classifier on test set: 0.9264705882352942
```

4) Fail case images (unsorted and all)









5) Fail Case Images (Sorted and limit to 2 per class)

