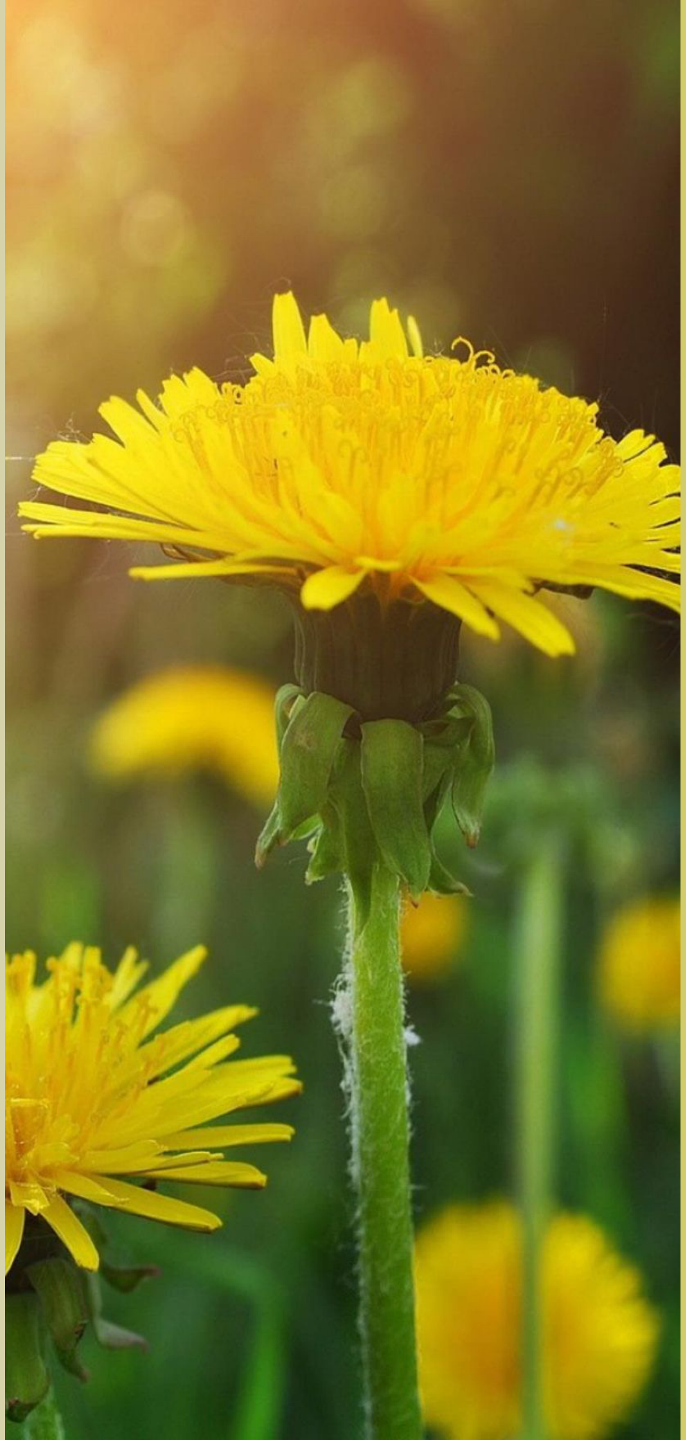
CLASSIFYING WILD EDIBLE FLOWERS BY COLOR SEGMENTATION AND HISTOGRAM OF ORIENTED GRADIENTS

By Cody Le, DePaul University, CSC481, Fall 2021



INTRODUCTION:

- Previous Studies
- Objective/Purpose
- Images and Dataset

PREVIOUS STUDIES

- Agarwal, Sonali et. al (2018): Framework for plant leaf recognition. Classified leaves with natural background using Support Vector Machines (SVM).
- Ibrahim, Zaidah et. al (2018): Compared texture features for classifying herbal plant leaves. Determined that HOG feature extraction was best.
- Shaparia, Riddhi et. al (2017): Classified flowers using color and texture feature extracted by Gray Level Co-occurrence Matrix (GLCM) through neutral network classifier (5 classes).
- Galphade, Ashish and Walse K.H. (2019): Classified flowers using color, GLCM, and shape with Artificial Neutral Network (ANN) (4 classes).
- Hoang, Thai et. al (2013): Proposed a multi-class SVM model to classifying flowers using color features (3 classes).

OBJECTIVE AND PURPOSE

Objective:

- Classify flowers based on images of flower head and petals with various backgrounds
- Use color segmentation to segment images prior to feature extraction
- Perform classification using Histogram of Gradients (HOG) feature extraction

Goals/Purpose:

- Compare classifier model with different number of flower classes
  - 5 classes: Different color and shaped flowers, then
  - 7 classes: Introduce some similar colored flowers
  - 9 classes: Introduce similar color and shaped flowers
- Determine if proposed methodology can classify flowers with greater number of classes and with visually similar looking objects.

DATA AND IMAGE SELECTION

Wild Edible Plants
Dataset from Kaggle

- 35 types of plants
- Images obtained through Flickr

Image Selection

- 9 types of flowering plants
- 25 images of each flower
- Only images with full flower head and flower petals selected.
- Images contained various backgrounds: blurred, greenery, natural, bright, dark, shadows, in various lighting.

SELECTED FLOWERS FOR ANALYSIS

Yellow Flowers:

- Dandelion
- Calendula
- Coltsfoot

White Flowers:

- Daisy
- Gardenia

Violet/Pink Flowers:

- Geranium
- Knapweed
- Chicory
- Common Mallow

METHODOLOGY:

1. Image Preprocessing:
   - Color Segmentation
2. Create Training and Test Sets
3. HOG Feature Extraction
4. Classification with SVM
5. Evaluate and compare with higher number of classes.
   - With Similar Color
   - With Similar Color & Shape

COLOR SEGMENTATION

All images contain a flower and various backgrounds. To separate the flower from the background, color segmentation was applied using the Red or Blue color channel that more closely aligned with the color of the flower.

Images are then converted to binary, cleaned, and masked to display segmented image.

Flowers using Red Color Channel for Segmentation:

- Daisy (White/Yellow Colored Flower)
- Dandelion (Yellow Colored Flower)
- Calendula (Yellow/Orange Colored Flower)
- Coltsfoot (Yellow Flower)

Flowers using Blue Color Channel for Segmentation:

- Gardenia (White Colored Flower)
- Geranium (Violet Colored Flower)
- Chicory (Violet Colored Flower)
- Knapweed (Pink Colored Flower)
- Common Mallow (Pink Colored Flower)

IMAGE PREPROCESSING: SEGMENTATION PROCESS

1. All images were cropped to have equal length and width (square image)
2. Convert image to grayscale using red or blue color channel
3. Convert grayscale image to binary using Otsu's method
4. Fill in regions/holes in binary image
5. Perform area opening to remove small objects in binary image
6. Mask original image to cleaned binary image
7. Resize image to 400 by 400 pixels using Nearest-Neighbor Interpolation

Perform in Bulk Using a Loop in MATLAB:

```
Images = dir('C:\folder_pathway\*.jpg')
outDir = 'C\new_folder_pathway\';
mkdir(OutDir);
for i = 1:length(Images)
   ImgName = strcat('C:\folder_pathway\', Images(i).name);
   Img = imread(ImgName);
   grayImg = Img(:, :, 1);
   bw = imbinarize(grayImg);
   binImg = imfill(bw, 'holes');
   binImg = bwareaopen(binImg, 400);
   maskedImg = bsxfun(@times, Img, cast(binImg, class(Img)));
   newImg = imresize(maskedImg, [400 400], 'nearest');
   imwrite(newImg, strcat(outDir, Images(i).name));
end
```

VISUALIZATION OF SEGMENTATION PROCESS

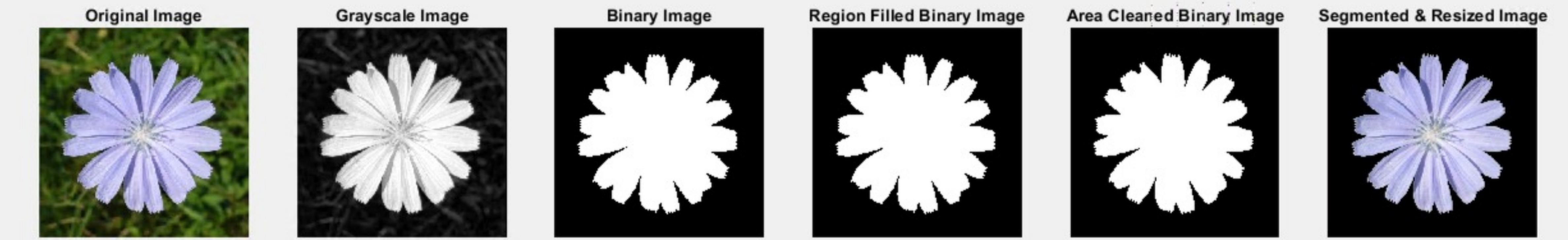
DAISY:

CHICORY:

VISUALIZATION OF SEGMENTATION PROCESS

DAISY: Using Red Color Channel

CHICORY: Using Blue Color Channel



TRAINING & TESTING SETS

Training and Testing Sets:

- Randomized Split
- 80 / 20 Split

  (80% Training, 20% Testing)
- Training Set: 20 images
- Testing Set: 5 images
- Training and Testing Sets created for each separate analysis.

DATASETS INFORMATION

First Analysis: 5 Types of Flowers

- Daisy, Dandelion, Gardenia, Geranium, Knapweed
- 2 Red, 3 Blue Color Channeled Flowers
- Flowers with different color, texture and shape

2nd Analysis: 7 Types of Flowers

- Calendula, Chicory, Daisy, Dandelion, Gardenia, Geranium, Knapweed
- 3 Red, 4 Blue Color Channeled Flowers
- Flowers with similar colors (yellow and violet added)

3rd Analysis: 9 Types of Flowers

- Calendula, Chicory, Coltsfoot, Common Mallow, Daisy, Dandelion, Gardenia, Geranium, Knapweed
- 4 Red, 5 Blue Color Channeled Flowers
- Flowers with similar colors and shape/texture added

1ST ANALYSIS: 5 TYPE OF FLOWERS

Daisy – Dandelion - Gardenia – Geranium - Knapweed

2ND ANALYSIS: 7 TYPE OF FLOWERS

Calendula – Chicory - Daisy – Dandelion
Gardenia – Geranium - Knapweed

2ND ANALYSIS: 7 TYPE OF FLOWERS

Calendula – Chicory - Daisy – Dandelion
Gardenia – Geranium - Knapweed

3RD ANALYSIS: 9 TYPE OF FLOWERS

Calendula – Chicory – Coltsfoot
Common Mallow – Daisy – Dandelion
Gardenia – Geranium – Knapweed

3RD ANALYSIS: 9 TYPE OF FLOWERS

Calendula – Chicory – Coltsfoot
Common Mallow – Daisy – Dandelion
Gardenia – Geranium – Knapweed

FEATURE EXTRACTION: HISTOGRAM OF GRADIENTS(HOG)

Histogram of Gradients (HOG):

- Feature descriptor that counts occurrences of gradient orientation in a localized portion of an image.
- Similar approach to Edge Orientation Histograms
- HOG focuses on the structure or shape of the object
- Considered better than all edge descriptors as it uses magnitude as well as angle of the gradient to compute features.
- Determined by Ibrahim, Zaidah et. al (2018) in their research of herbal plant classification as best texture feature extraction approach.

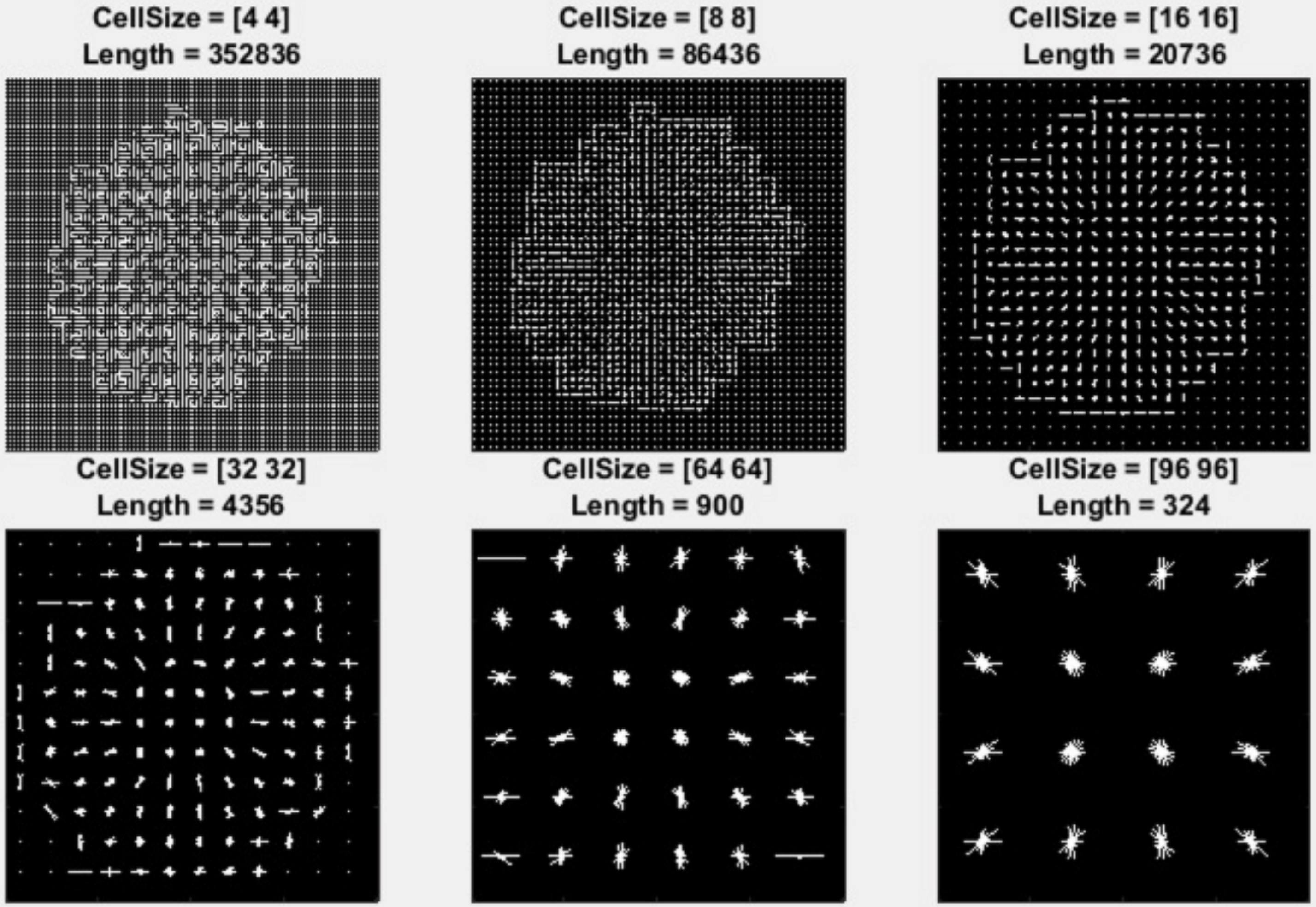
Calculating HOG Features using MATLAB:

- Determine HOG Feature Size
- extractHOGFeatures function returns a visualization output to evaluate the right amount of information
- Take sample image from training set, extract HOG features and plot the visualization
- From visualization plots, determine which size parameter encodes best amount of shape information
- Best HOG Feature Size should encode enough spatial information with the lowest dimensionality

VISUALIZATION OF HOG EXTRACTED FEATURES

Image from Training to Determine HOG Size:

64 x 64: 900 Features
32 x 31: 4356 Features
96 x 96: 324 Features



Analysis will test 32x32, 64x64, and 96x96 with classifier to determine which HOG Size performed best.

CLASSIFICATION WITH SUPPORT VECTOR MACHINES (SVM)

- Multi-class Support Vector Machines (SVM) is used for the classification.
- Agarwal, Sonali, et. Al (2018) have stated in their research that a multi-class SVM provides the best accuracy for plant recognition with multiple classes.
- SVM are supervised learning models for classification.
- Maps training samples to points in space to maximize the boundaries between classes
- Test samples are then mapped into the same space and predicted to belong to a class based on which side of the boundary they fall in.
- ’One-to-One’ Encoding Scheme using MATLAB fitcecoc function
  - Splits multi-class classification into binary classification
  - Splits points based on each class
  - Most popular and affective approach for SVM

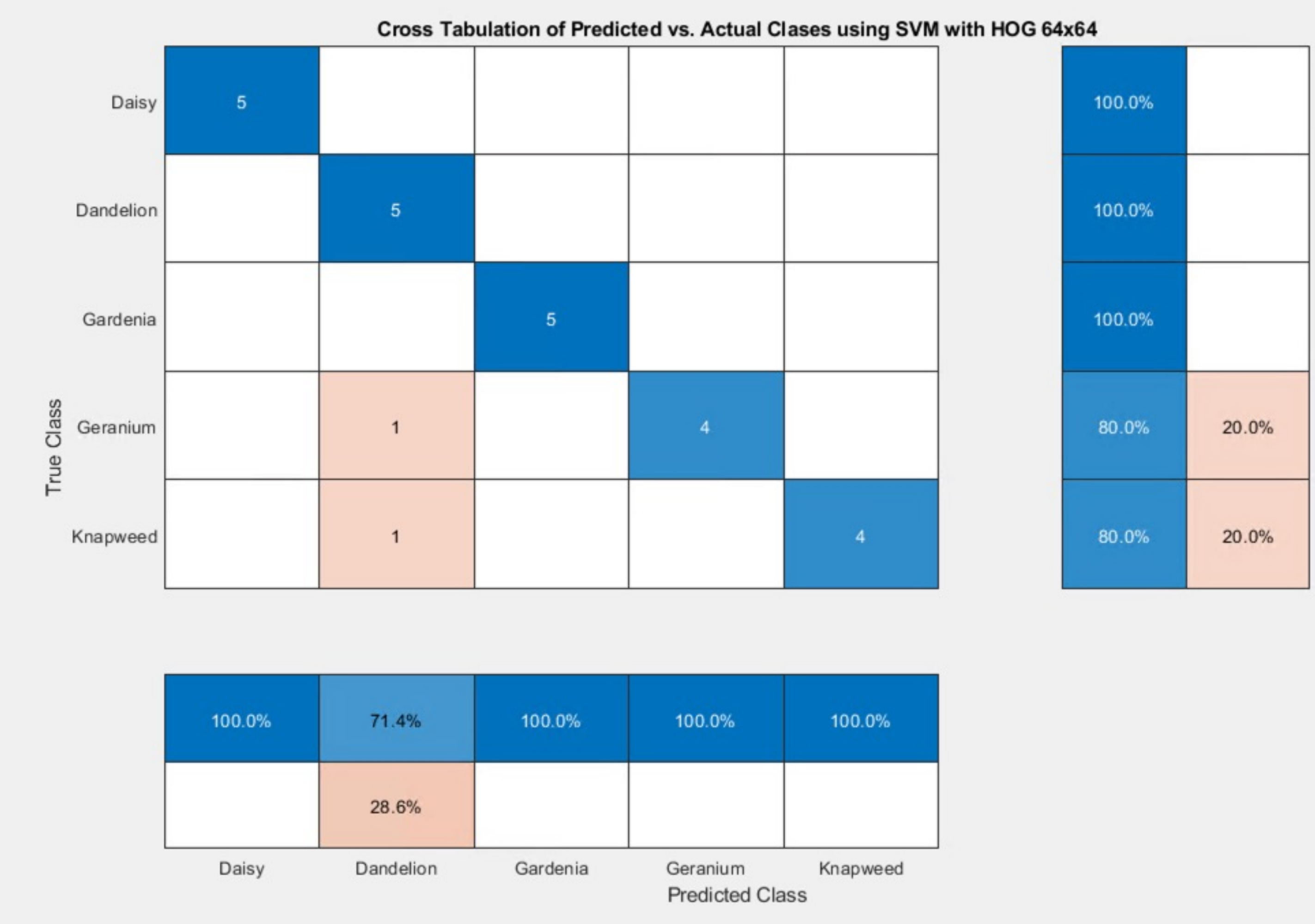
SVM MODEL AND EVALUATION PROCESS

1. Extract HOG Features from Training Set
2. Set Training Set Labels
3. Fit SVM Model using Training Set Features and Training Set Labels
4. Extract HOG Features from Testing Set
5. Set Testing Set Labels
6. Use SVM Model to Predict on Testing Set Features
7. Calculate Confusion Matrix
   (Test Labels vs. Predicted Labels)
8. Evaluate Results Using
   K-Fold Cross-Validation using MATLAB function, crossval

RESULTS:

- 1st Analysis: 5 Types of Flowers
- 2nd Analysis: 7 Types of Flowers
- 3rd Analysis: 9 Types of Flowers
- Compare HOG Feature Size
- Compare SVM Model Accuracy

RESULTS OF 1ST ANALYSIS: 5 TYPE OF FLOWERS



Best HOG Size: 64x64

Accuracy from Cross-Tab:
92%

Estimated Loss Using
Cross-Validation:
11%

Estimated Accuracy with
Cross-Validation:
89%

RESULTS OF 1ST ANALYSIS: HOG COMPARISON

5 Classes of Flowers:

HOG Feature Size	Model Accuracy	Estimated Error with Cross-Validation	Estimated Cross-Validation Accuracy
64 x 64	92%	11%	89%
96 x 96	88%	17%	83%
32 x 32	84%	10%	90%

- **64 x 64 provided the best performance with model accuracy of 92%**
- **With lower features (96 x 96), estimated error increases.**
- **With higher features (32 x 32) estimated error slightly decreases.**

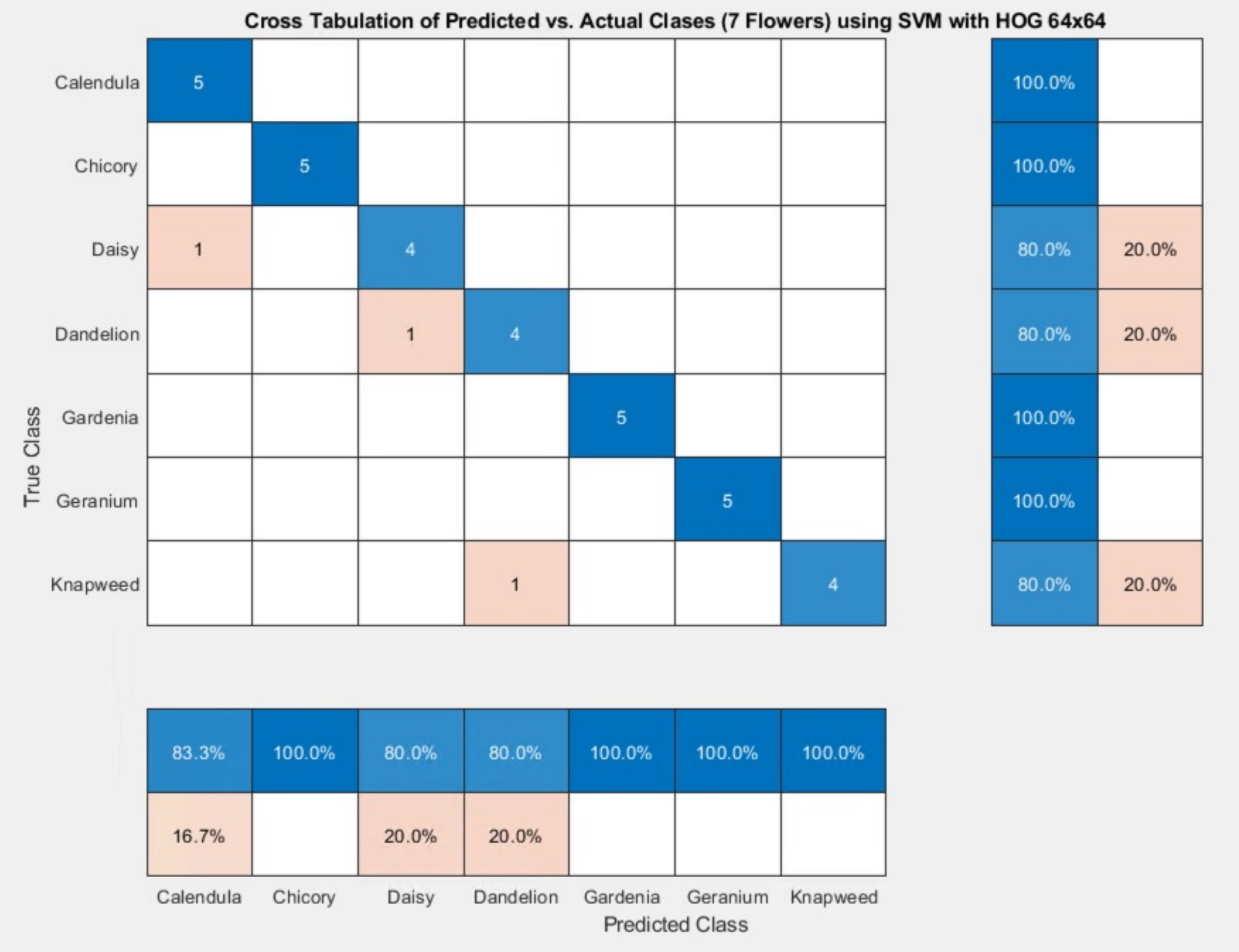
RESULTS OF 2ND ANALYSIS: 7 TYPE OF FLOWERS

Best HOG Size: 64x64

Accuracy from Cross-Tab: 91.43%

Estimated Loss Using Cross-Validation: 22.86%

Estimated Accuracy with Cross-Validation: 77.14%



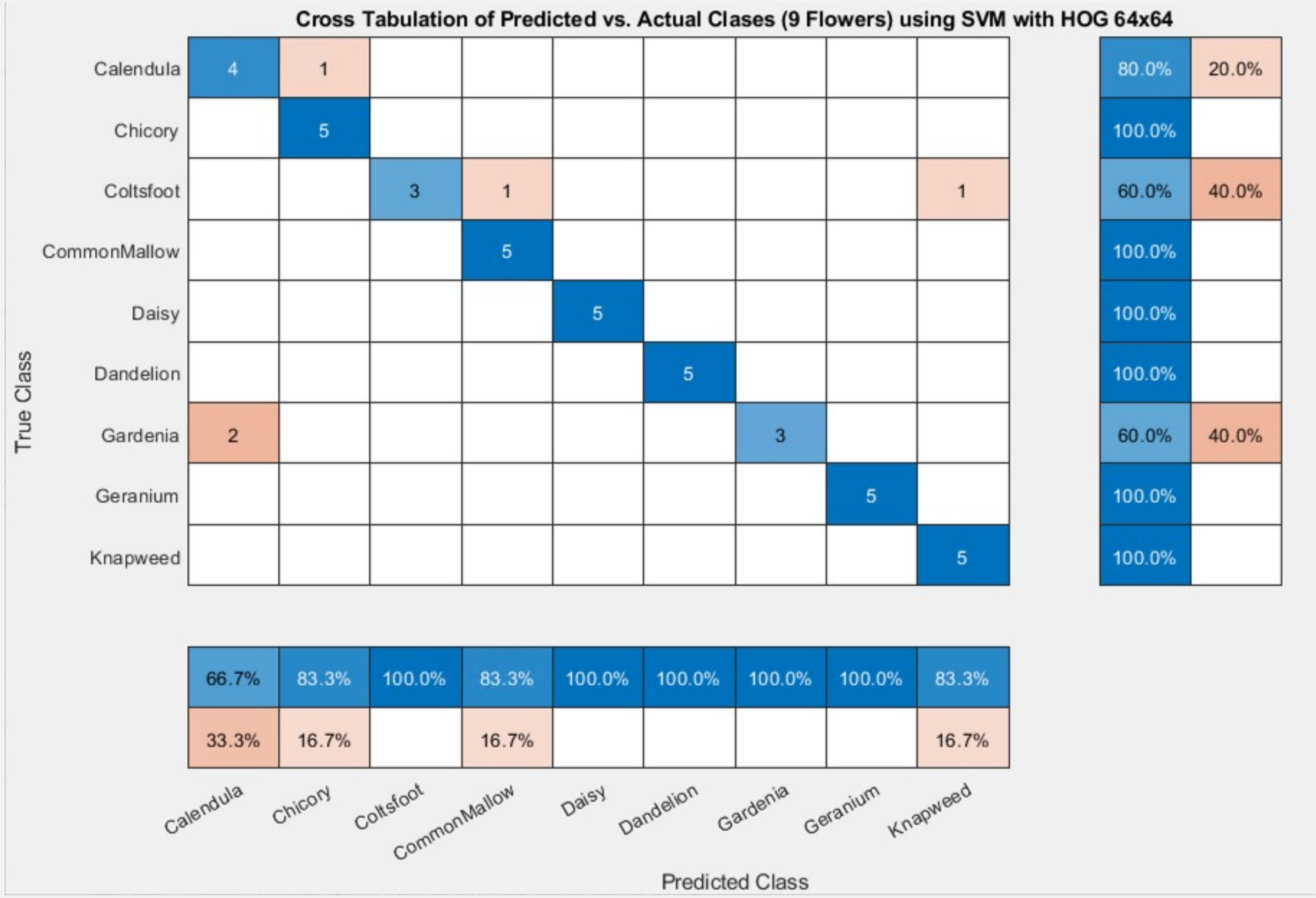
RESULTS OF 2ND ANALYSIS: HOG COMPARISON

7 Classes of Flowers:

HOG Feature Size	Model Accuracy	Estimated Error with Cross-Validation	Estimated Cross-Validation Accuracy
64 x 64	91.43%	22.86%	77.14%
96 x 96	82.86%	30%	70%
32 x 32	82.86%	30.71%	69.29%

- **64 x 64 provided the best performance with model accuracy of 91.43%**
- **Estimated error is higher with 7 classes.**
- **96 x 96 and 32 x 32 perform equally lower than 64 x 64 with estimated loss much higher than 64 x 64.**

RESULTS OF 3RD ANALYSIS: 9 TYPE OF FLOWERS



Best HOG Size: 64x64

Accuracy from Cross-Tab: 88.89%

Estimated Loss Using Cross-Validation: 31.11%

Estimated Accuracy with Cross-Validation: 68.89%

RESULTS OF 3RD ANALYSIS: HOG COMPARISON

9 Classes of Flowers:

HOG Feature Size	Model Accuracy	Estimated Error with Cross-Validation	Estimated Cross-Validation Accuracy
64 x 64	88.89%	31.11%	68.89%
32 x 32	80%	33.89%	66.11%
96 x 96	77.78%	36.11%	63.89%

- **64 x 64 provided the best performance with model accuracy of 89.89%**
- **Estimated error is highest with 9 classes.**
- **32 x 32 performed better than 96 x 96.**

COMPARISON OF CLASSIFIER MODELS:

Model	Accuracy	Estimated Error with Cross-Validation	Estimated Cross-Validation Accuracy
5-Class w/ HOG 64x64	92%	11%	89%
7-Class w/ HOG 64x64	91.43%	22.86%	77.14%
9-Class w/ HOG 64x64	88.89%	31.11%	68.89%

- **5 Class Model performed best with the highest accuracy and the lowest estimated error.**
- **7 Class Model still performed well but had a higher estimated error compared to 5-class model.**
- **9 Class Model had the lowest accuracy and the highest estimated loss compared to the 5-class and 7-class model.**

CONCLUSION:

- Color Segmentation and HOG can be used to classify flowers.
- Best HOG feature size contains enough spatial information with lower dimensionality (in this case, 64 by 64).
- SVM model can predict multi-class flower classification using HOG with positive results.
- Accuracy lowers with increased class size and with greater number of similar flowers.
- Model performed well in classifying similar colored and shaped flowers but with a higher estimated error.

LIMITATIONS / DISCUSSION:

- Relatively low sample size of images.
- Selected flowers respond well to either a Red or Blue Color Channel
- Images had mostly blurred, dark, or green backgrounds such as greenery or nature.
- Most images are close-up shots of flowers only.

Future Explorations:

- Larger sample size of flower images.
- Compare larger number of similar colored or similar shaped flowers.
- Compare flowers during different growths (flower bud vs. flower opened vs. flower wilt).