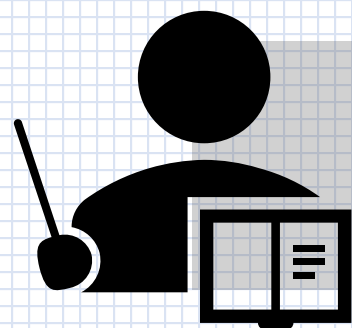


MACHINE LEARNING:

feature extraction



Professor: Dr. Maricor Soriano
Instructor: Jayson Cubero

Rene L. Principe Jr.
2015-04622



dataset

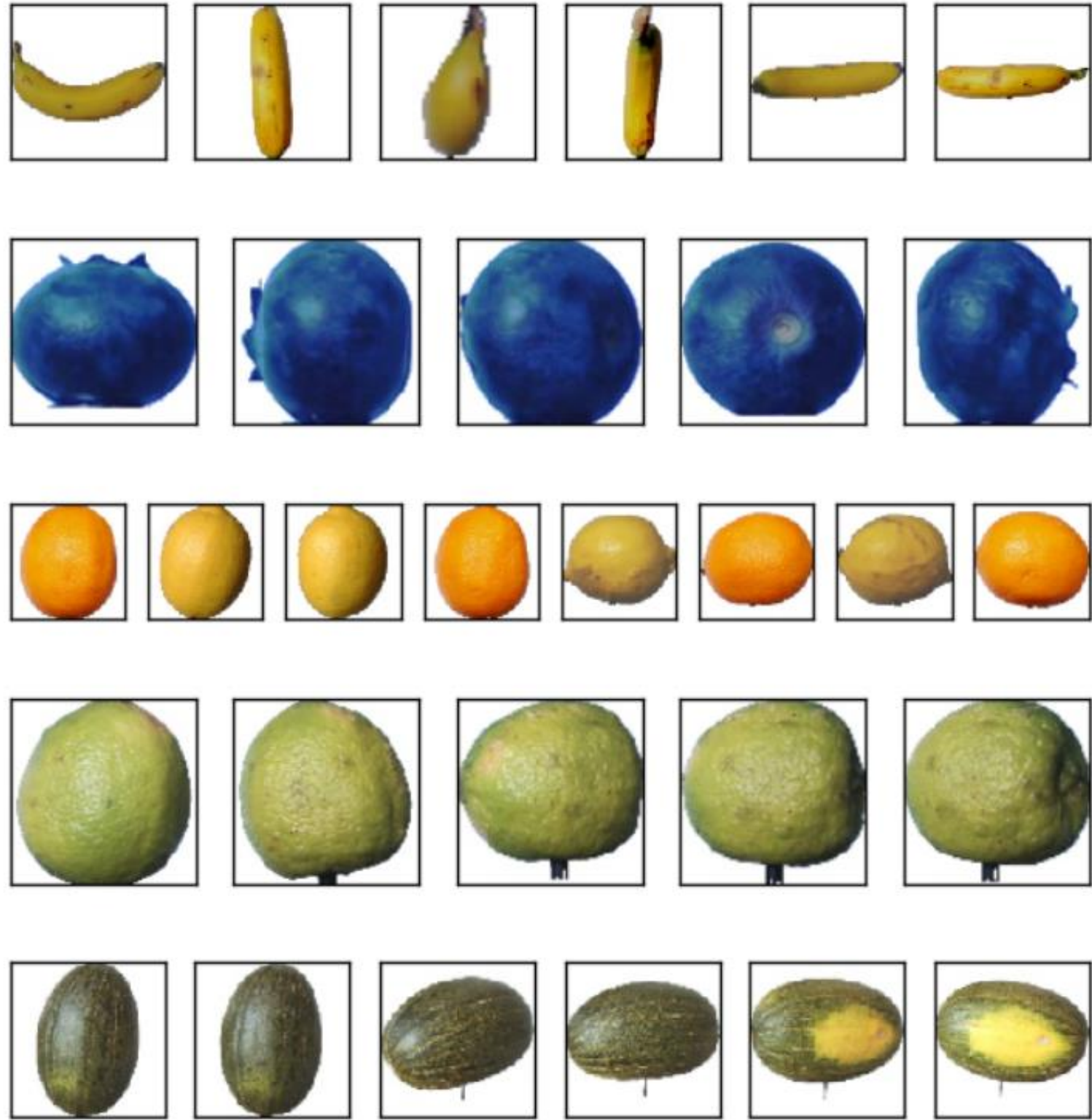


Figure 1. Images taken from Kaggle Fruits 360 dataset.

Feature extraction, an essential pre-processing step in Machine Learning, shall be demonstrated in this activity [1]. A readily available dataset from Kaggle of various fruits shot in 360 degrees was used. The homogeneity of the dataset was exploited for faster processing time. All 20,701 images of 38 different fruits were 100x100 pixels and in white background.

I aim to extract the color and eccentricity features of the 5 fruits I selected, namely: banana (5 images), blueberry (5 images), melon (6 images), lemon (8 images), and limes (5 images). The images I used are shown in Figure 1. Visually, the colors are discernible from one another. The high eccentricity of banana and melon is prominent. The goal of this activity is to automate this feature extraction and compare the fruits' quantified colors and eccentricity using the image processing skills I've learned previously.

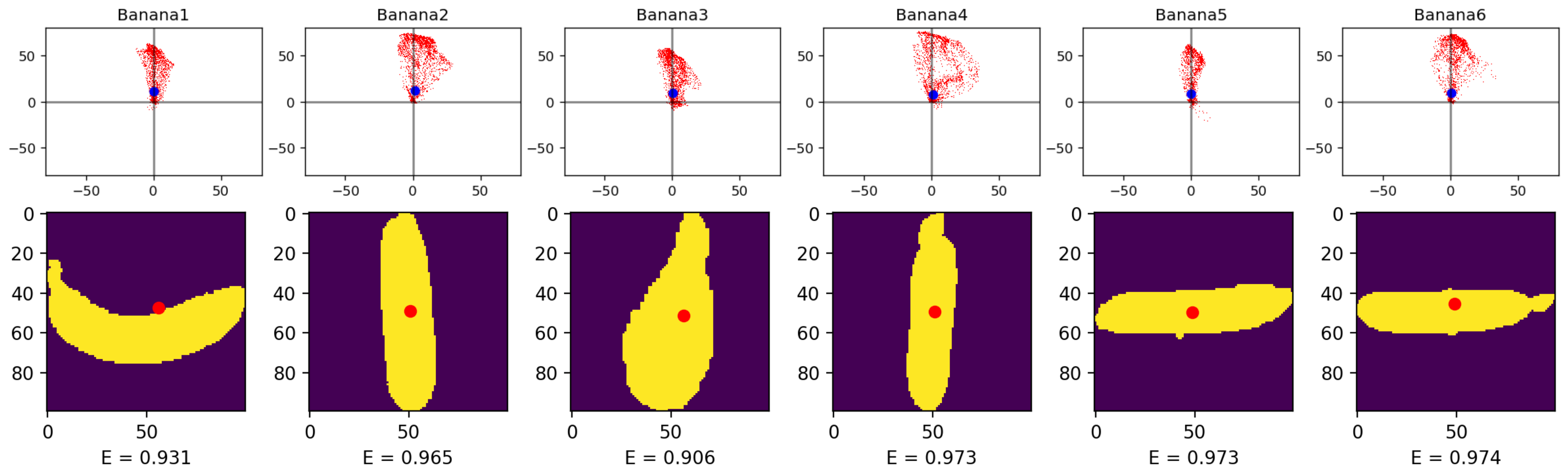


Figure 2. (Top row) Color Gamut of the fruit image in CIEL*a*b* colorspace. (Bottom row) Thresholding, segmentation, and blob analysis was implemented to measure the eccentricity E .

In this step, the images were loaded on a function which aims to return the mean a^* and b^* which represents the color feature and another function which uses `skimage.measure` to measure the eccentricity feature. Shown in Fig. 2 is the feature extraction step implemented on the Banana class. The yellow pixels due to the color of the banana yielded scattered points skewed towards positive a^* (y-axis) values in the color gamut. Meanwhile, high eccentricity implying that the banana is highly elliptical was measured using blob analysis. The values of E were consistently above 0.9. The eccentricity formula I used returns values near 0.5 when the blob is circular and will increase/decrease once the shape becomes more elliptical.

For all classes, the mean a^* and b^* and eccentricities are shown in Fig 3. Fruit colors were distinct in the LAB colorspace. Blue berry for example is located on the 4th quadrant near the $-a^*$ axis, representing bluish color of the fruit. Eccentricities of melon is around 0.75 while the banana is near 1, while all other fruits dwell on eccentricity values near 0.5 (nearly circular). These features were all combined into a 3D plot for visualization as shown in Fig. 4 and clustering was observed in the feature space. Five fruit classes were successfully represented as set of three features.

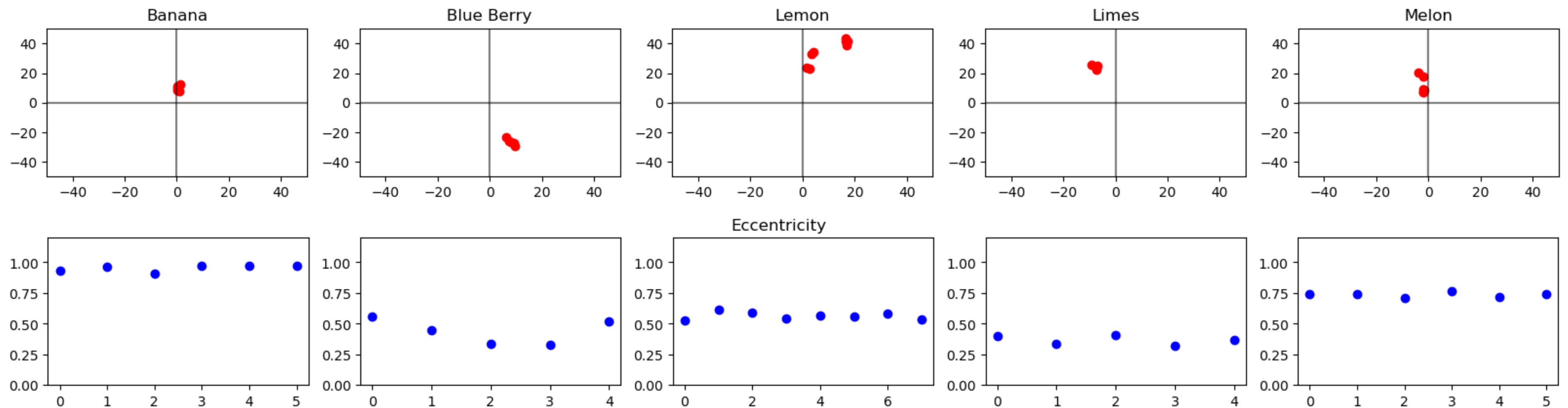


Figure 3. (Top row) Mean a^* and b^* of each class visualized in CIEL*a*b* colorspace. (Bottom row) Eccentricity of each fruit in each class. Homogeneous measured eccentricity is observed as manifested by nearly similar measures of E .

I have successfully extracted the color (a^* and b^*) features and shape (eccentricity) feature using image processing techniques and was able to show that five classes of fruits returned five clusters in the 3D feature space.

In this activity, I'd give myself an 11.

References:

- [1] M. Soriano, "Feature Extraction", 2019.
- [2] M. Oltean and H. Muresan, Fruits 360 dataset on kaggle. [Online; accessed 29.10.2019].

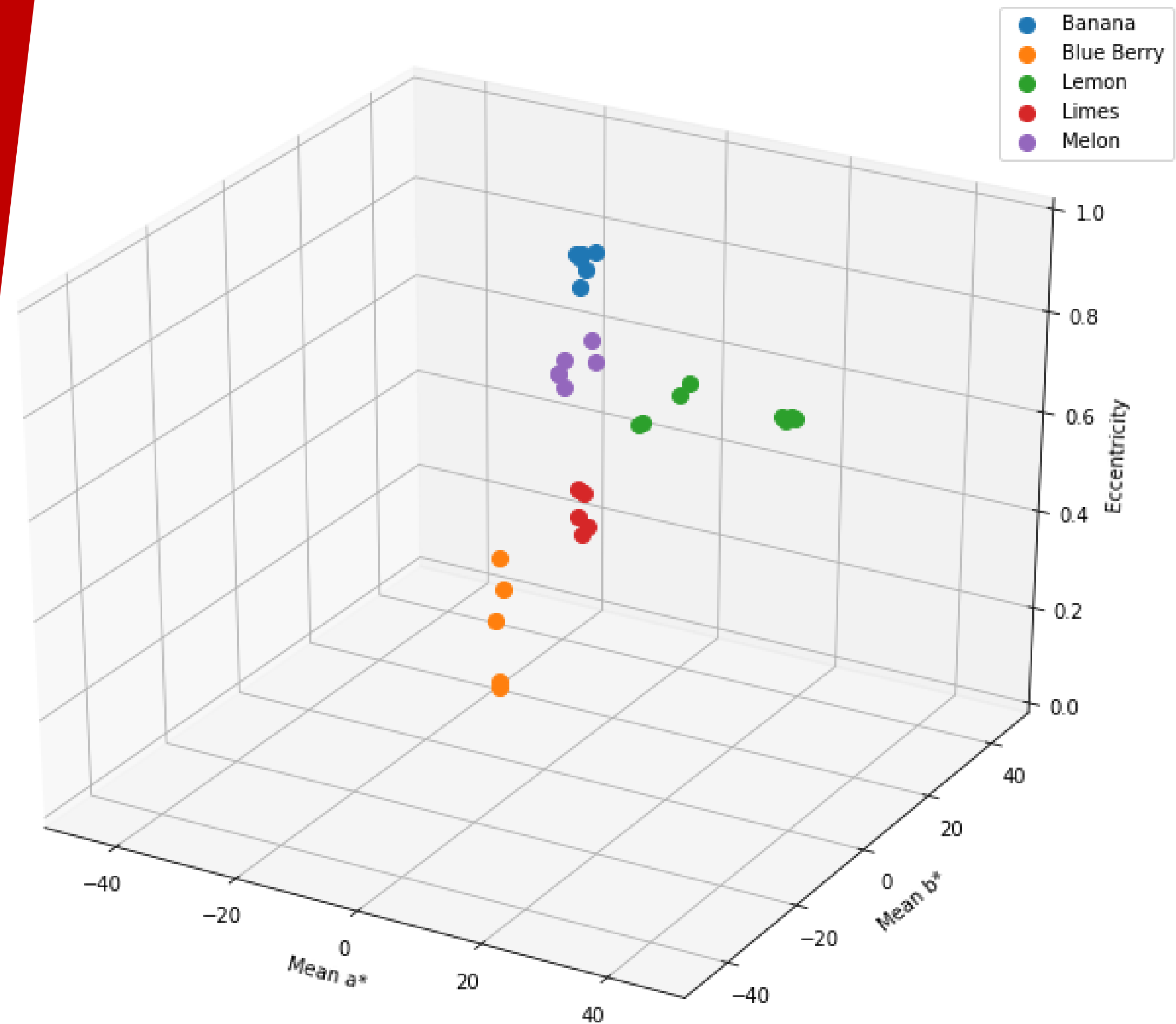


Figure 4. Combining the features, a 3D feature space was plotted above where a^* measures the greenness to redness, b^* measures the blueness to yellowness, and eccentricity measures how elliptical the fruit is.