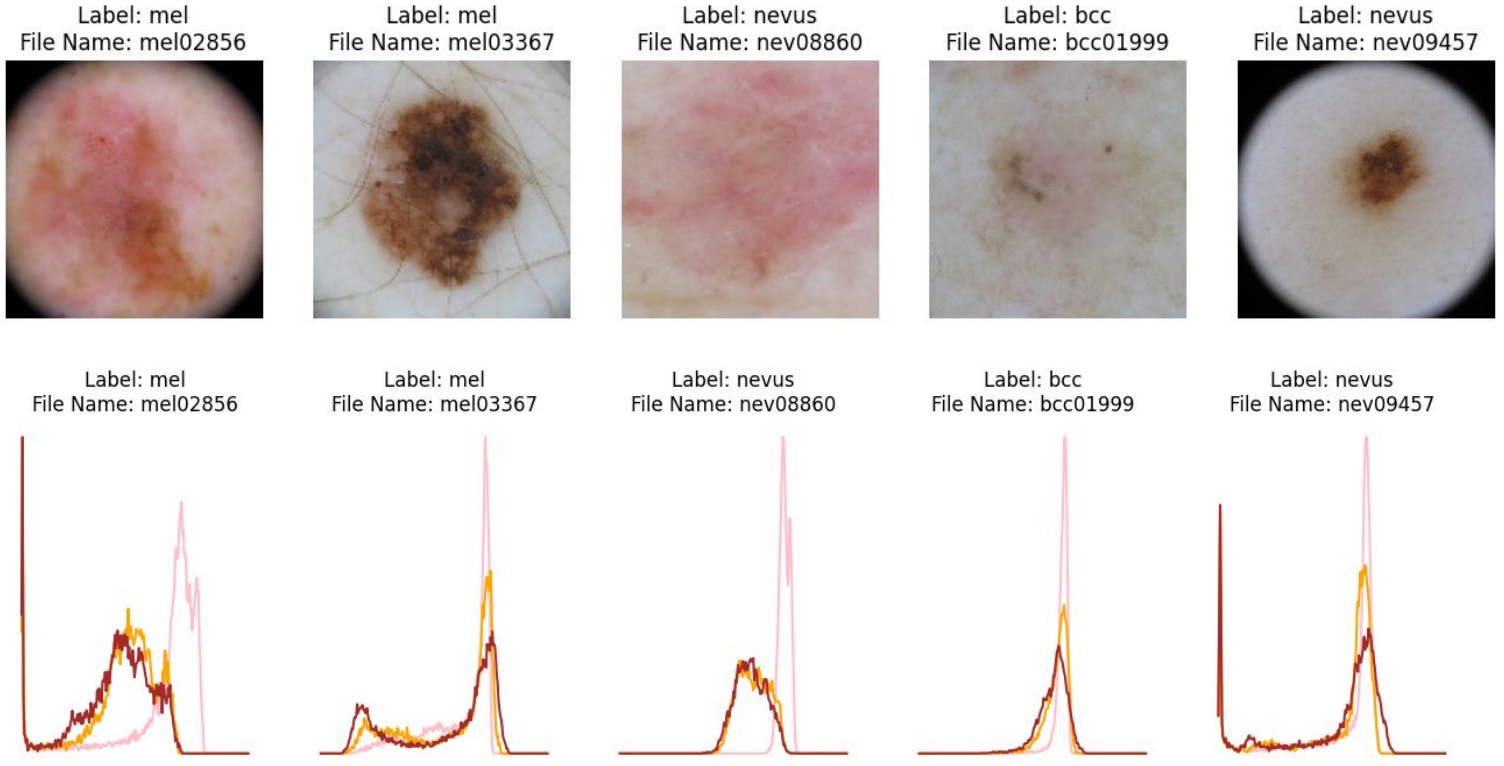
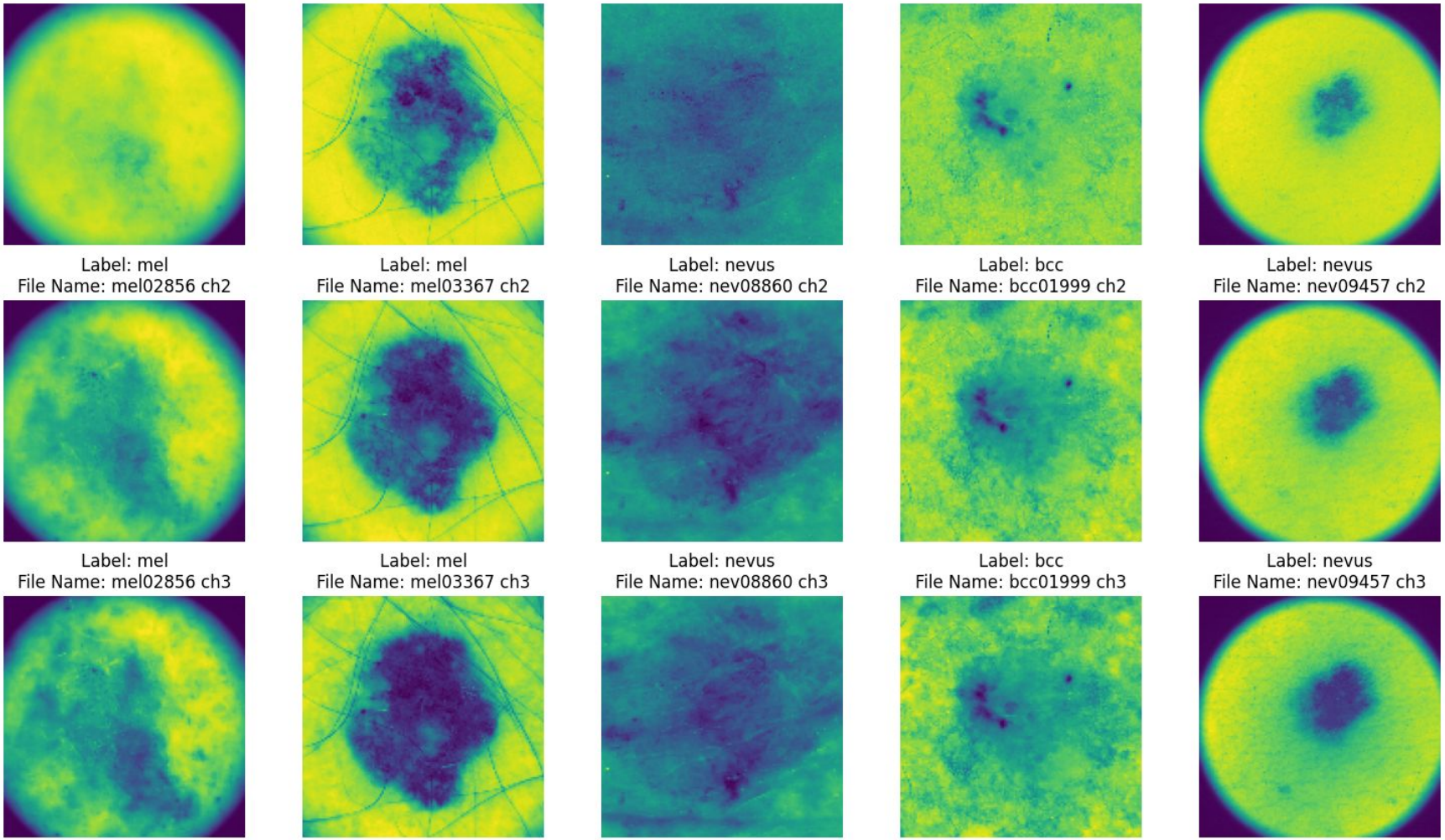
CAD Results

Edwing Ulin
Yusuf

First analysis of the data set




Label: mel
File Name: mel02856 ch1
Label: mel
File Name: mel03367 ch1
Label: nevus
File Name: nev08860 ch1
Label: bcc
File Name: bcc01999 ch1
Label: nevus
File Name: nev09457 ch1
Label: mel
File Name: mel02856 ch2
Label: mel
File Name: mel03367 ch2
Label: nevus
File Name: nev08860 ch2
Label: bcc
File Name: bcc01999 ch2
Label: nevus
File Name: nev09457 ch2
Label: mel
File Name: mel02856 ch3
Label: mel
File Name: mel03367 ch3
Label: nevus
File Name: nev08860 ch3
Label: bcc
File Name: bcc01999 ch3
Label: nevus
File Name: nev09457 ch3

Features - Color

color_statistics(image, color_space): Calculates first-order statistics and color channel correlations for an image in a specified color space ('RGB', 'HSV', 'LAB', or 'GRAY').

color_hist_bins(image, n_bins, color_space): Generates a histogram with a specified number of bins for each channel in the given color space.

Features - Texture

calculate_glcms(image, distances, angles, properties, color_space): This function computes GLCMs for an image at multiple distances and angles, then extracts specified texture properties for each GLCM.

- distances = [8,16,32],
- angles = [0,π/4,π/2,3π/4],
- properties = ['contrast','dissimilarity','homogeneity','energy','correlation','ASM'], color_space='HSV'

gloh_data(image, color_space): This function computes the mean, variance, and median of the histogram of a grayscale image for texture description.

Features - Shape

shape_measurements(image, color_space): Extracts shape descriptors from an image based on contours found within it.

Area: The total area covered by the contours.

Perimeter: The total length of the contours.

Dispersity: A measure of how spread out the contour is; it's the **square of the perimeter divided by the area**.

Saturation: Represents the **compactness** of the shape; it's the **area divided by the perimeter**.

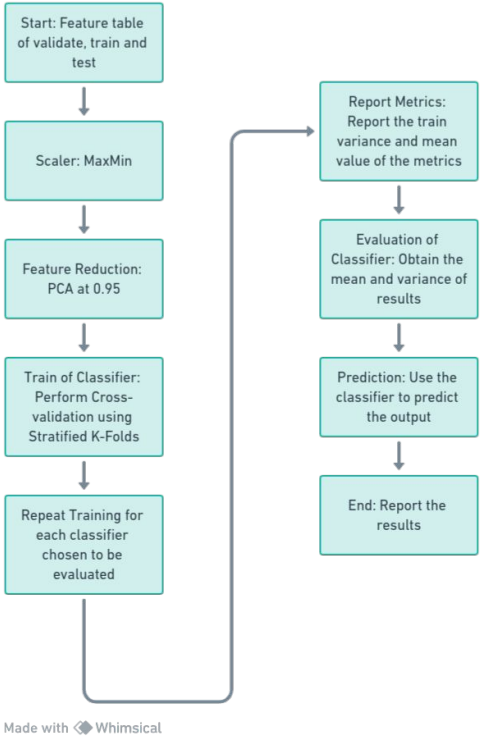
Roundness: A measure of how circular the contour is; calculated as **4π×Area divided by the square of the perimeter**.

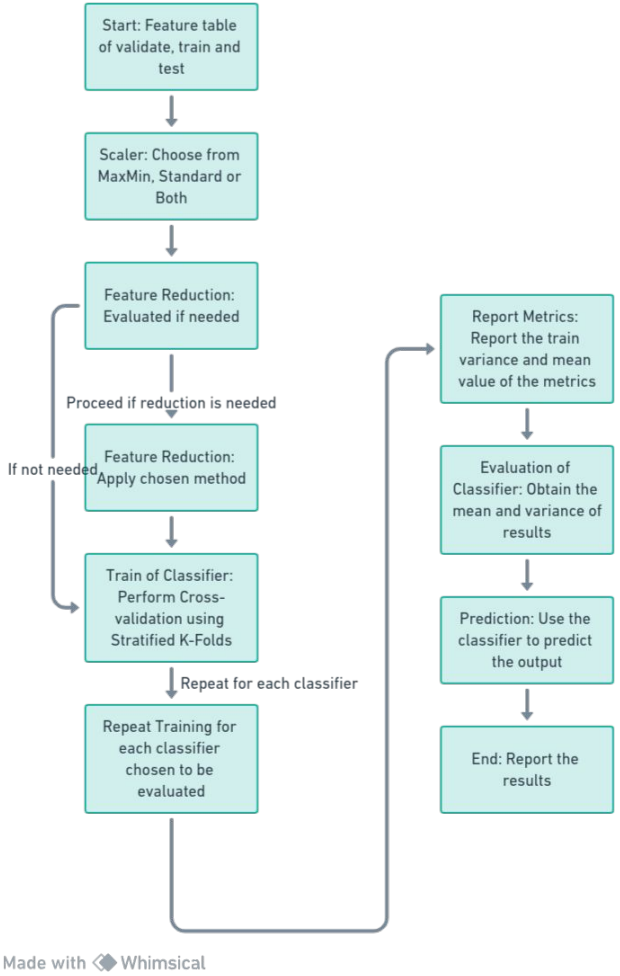
Pipeline for Binary Classification

From the features chosen:

- color.color_statistics, glcm.calculate_glcms, shape.shape_measurements, gloh.gloh_data

Drop the **color bins** due to noise.





Results of Binary

Color-Texture

Class	ACC mean	f1	recall
KN	0.68	0.68	0.38
RF	0.70	0.70	0.70
SVC	0.765	0.765	0.765

Color-Texture-Shape 125 Features

Class	ACC mean	f1	recall
SVC	0.771	0.770	0.771
SVC (gama auto)	0.746	0.745	0.746

Color-Texture-Shape 16 features

Class	ACC mean
SVC	0.777

Pipeline for Multi-Classification

- Normalization by scaling.
- Oversample minority class randomly.
- We use soft-voting classifier composed of
  - Decision Trees, Random Forest and a SVM.
- Fit using initial class weights to handle imbalance.

classes	mel	bcc	scc
class weights	0.624	0.890	4.505

Results of Multi-Classification

- We report validation performance of the model.

classes	mel	bcc	scc	total
Cohen-Kappa Scores	0.786	0.723	0.489	0.550

- We see the performance rise comparing to average performances of CV scores for two binary model.

Classifier	Accuracy	Cohen-Kappa	F1-Score	Recall
Random Forest Classifier	0.523	0.063	0.487	0.523
K-Neighbors Classifier	0.531	0.147	0.526	0.531

Thanks for your attention