CAD: Skin Lesion Classification

Manasi Kattel
Vladyslav Zalevskyi

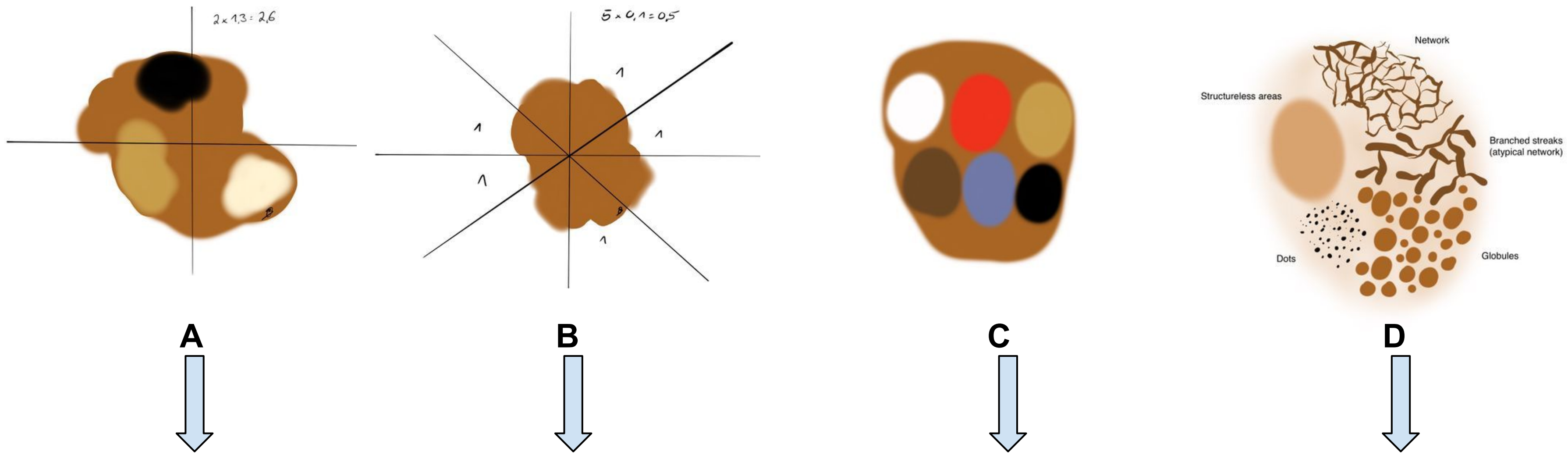
Universitat de Girona

Content

1. Literature review: ABCD rule
2. Preprocessing
   a. Hair removal
   b. Segmentation
3. Feature Extraction
   a. Color
      i. Color preprocessing
   b. Texture
   c. Shape
4. BoW
5. Challenge 1
   a. Feature selection/dimensionality reduction
   b. Results and experiments
6. Challenge 2
   a. Feature selection/dimensionality reduction
   b. Results and experiments
7. Conclusions

Universitat de Girona

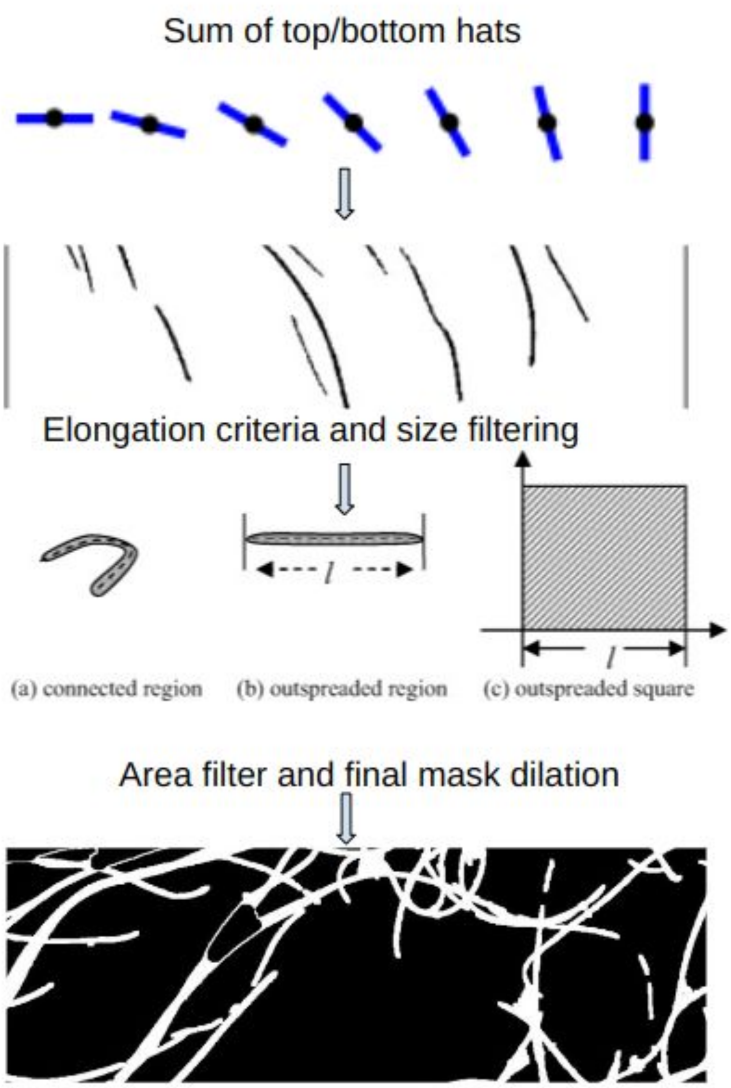
Literature Review: ABCD Rule

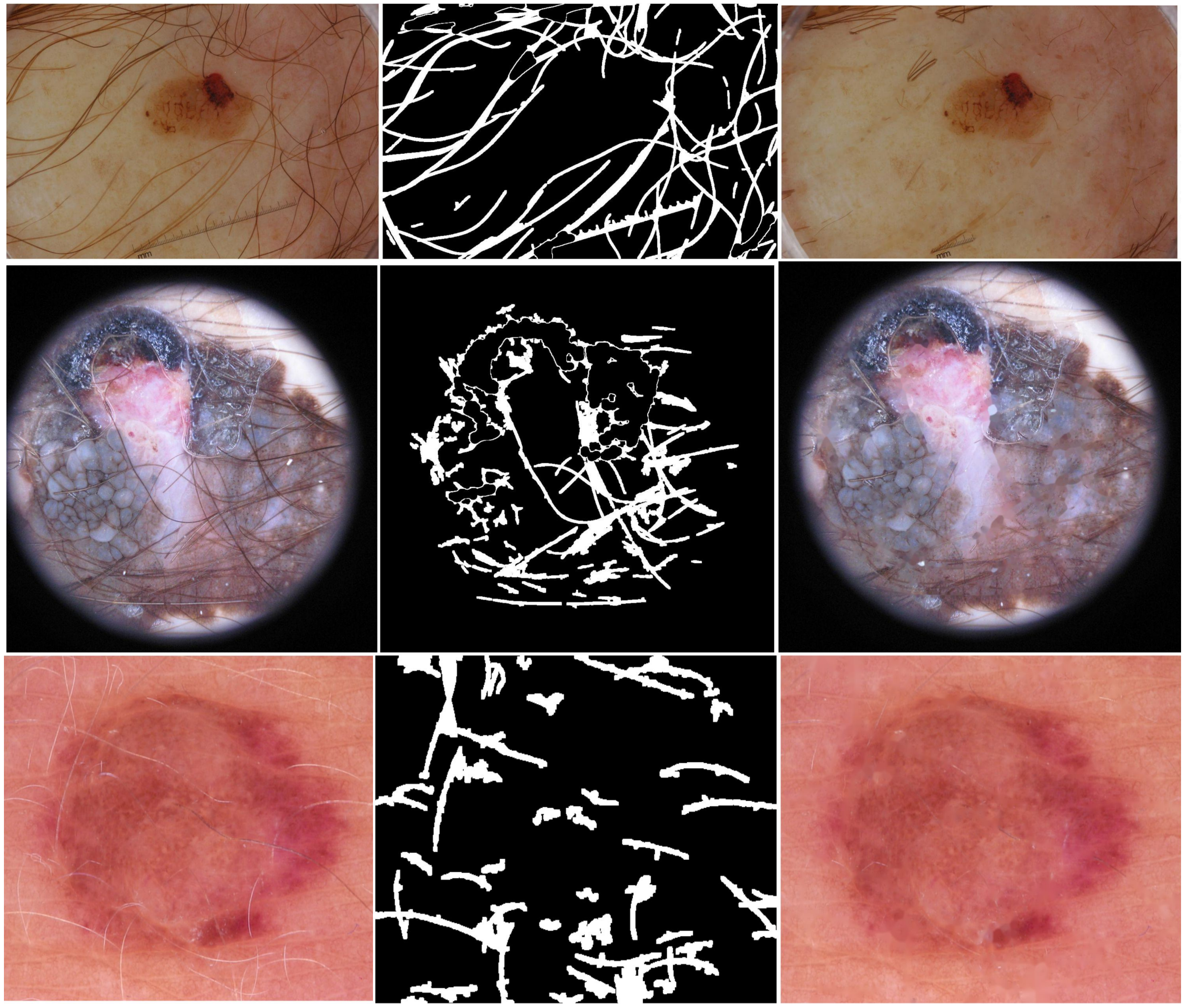


Asymmetry	**Border**	**Color**	**Dermoscopic structures**
• Perform segmentation • Extract features from the lesion mask • Highest weight	• Perform segmentation • Analyze textures and color at the lesion border • Lowest weight	• Transform to different color spaces • Extract features per channel • Try BoW	• Perform segmentation • Analyze textures and color at the lesion border • Try BoW

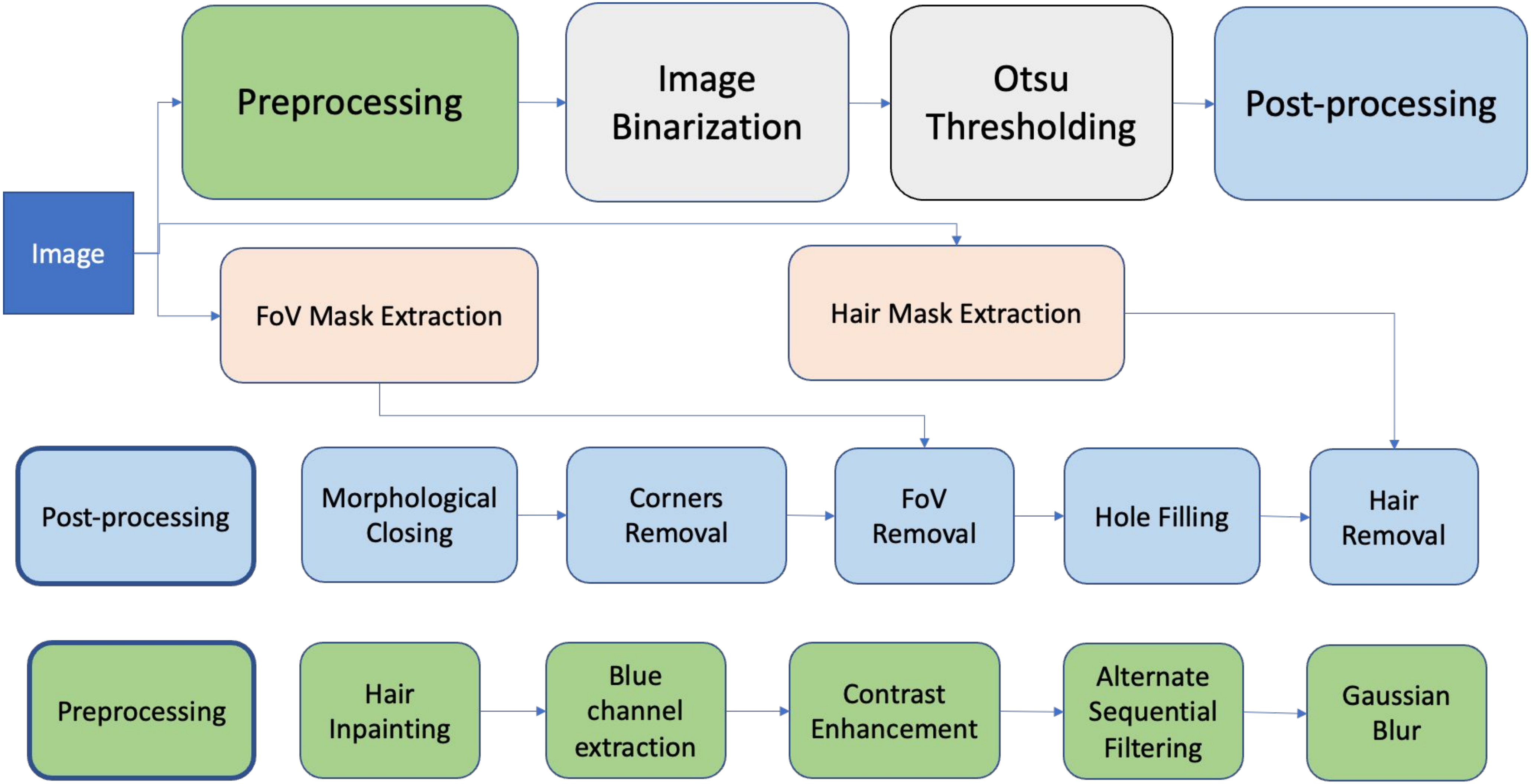
Universitat de Girona

Preprocessing: Hair Removal



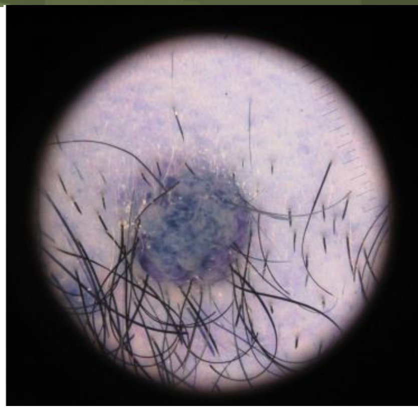


Preprocessing: Segmentation Pipeline



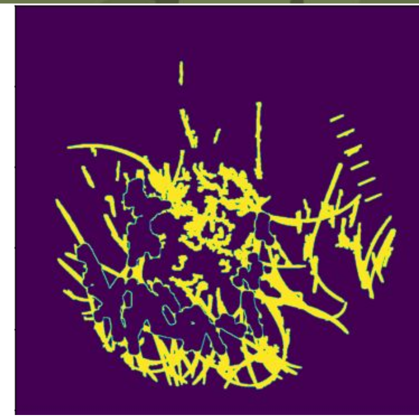
Universitat de Girona

Preprocessing: Segmentation Pipeline

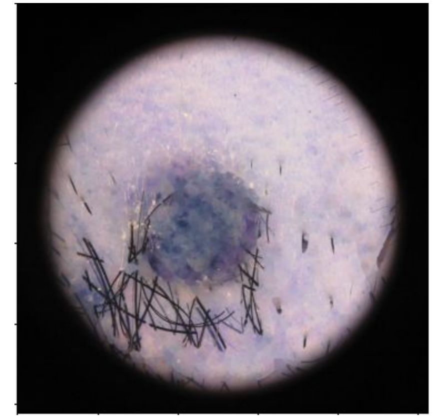


Original Image

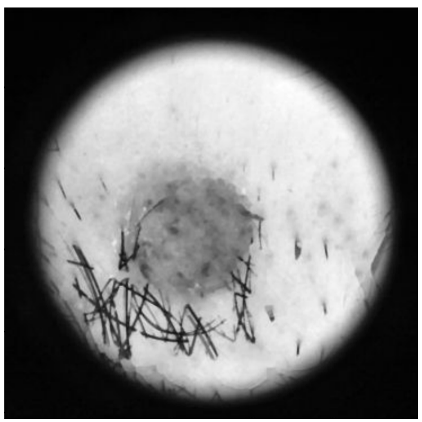
FoV Mask



Hair Mask

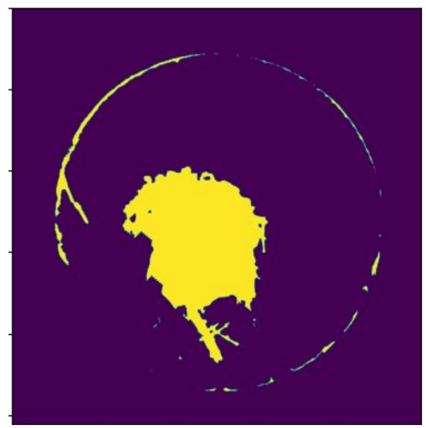


Inpainted

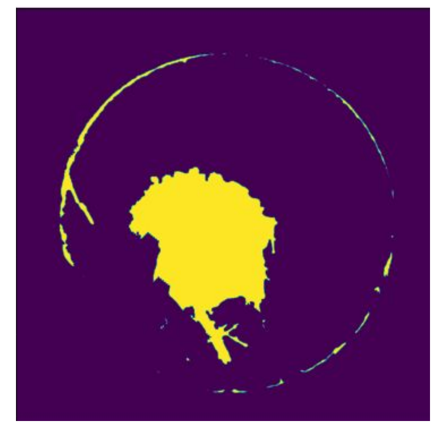


Enhanced and smoothed

Otsu Thresholding



Remove FoV



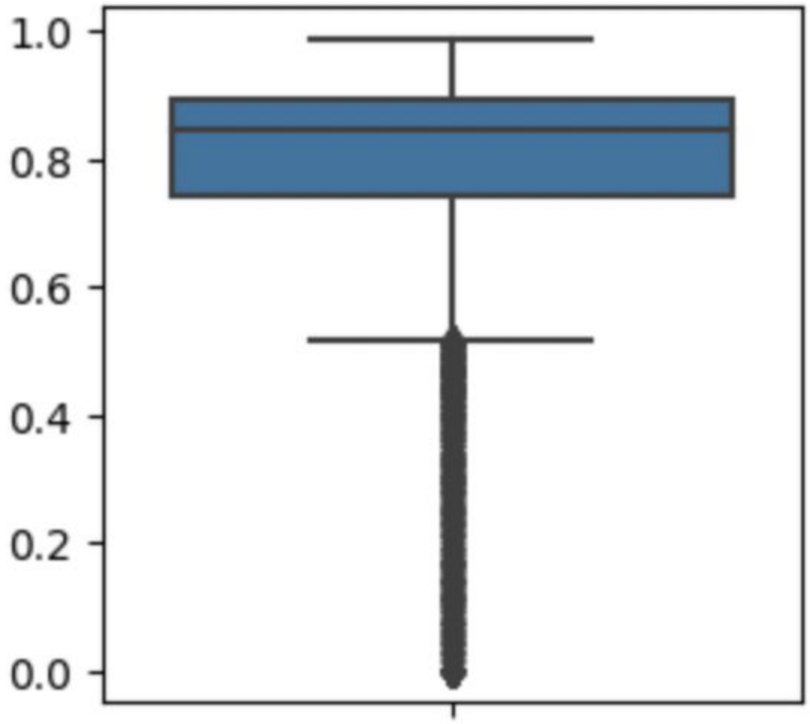
Fill Holes

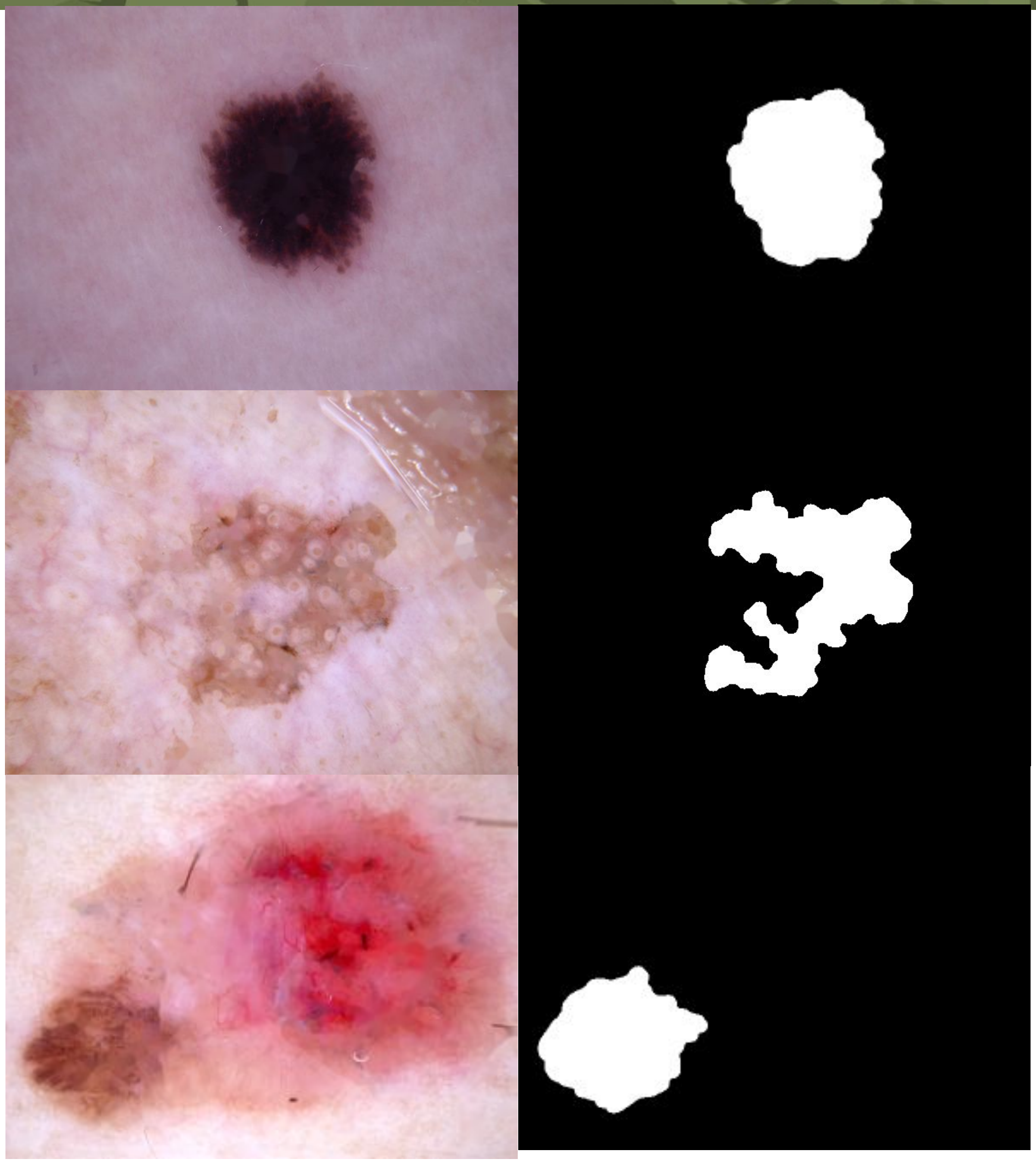
Final Segmentation Mask

Universitat de Girona

Preprocessing: Segmentation

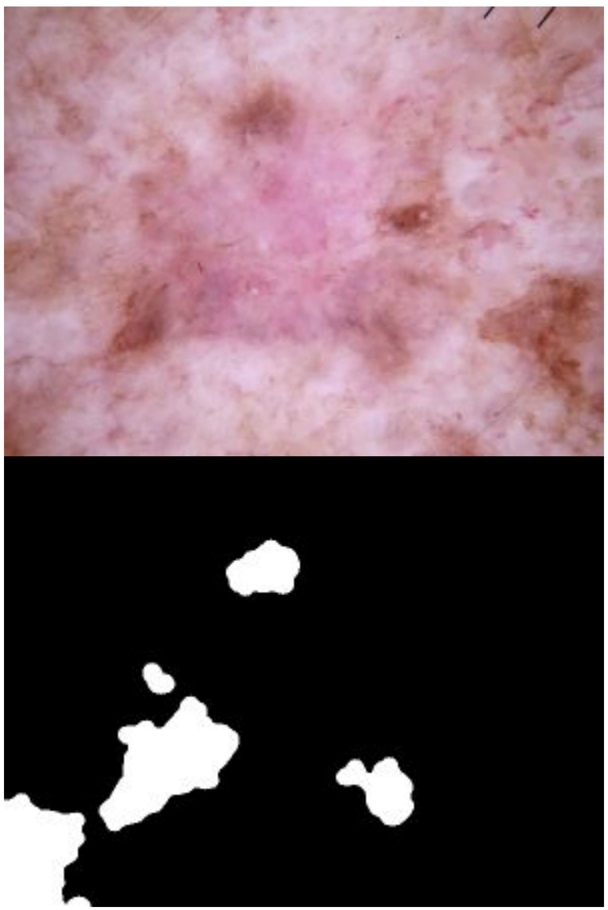
Dice scores of the developed segmentation algorithm reported on the HAM10000 dataset

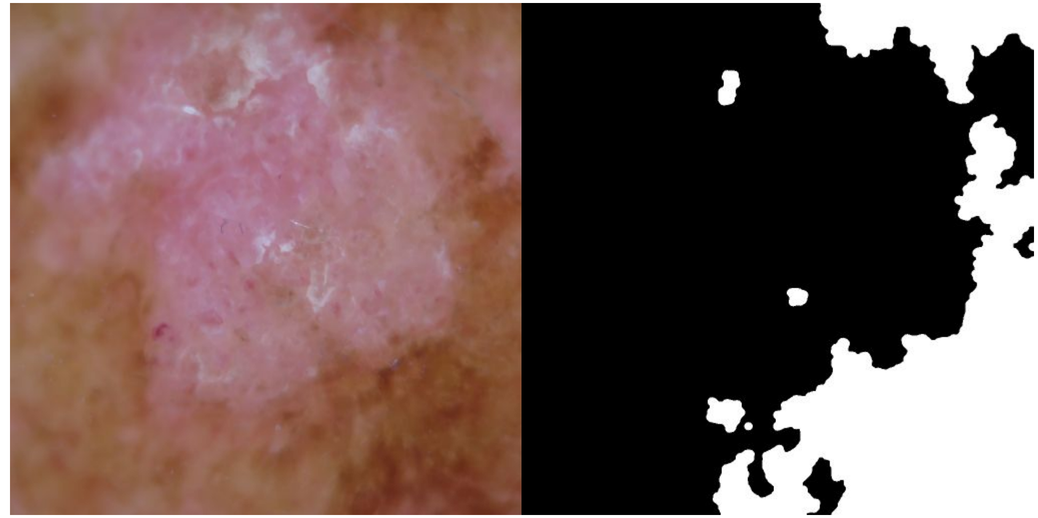


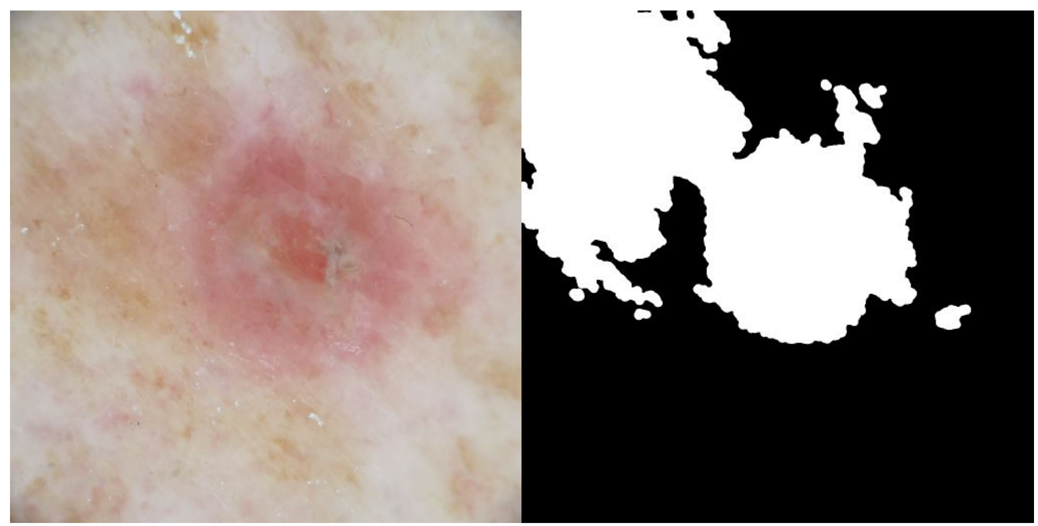


Segmentation

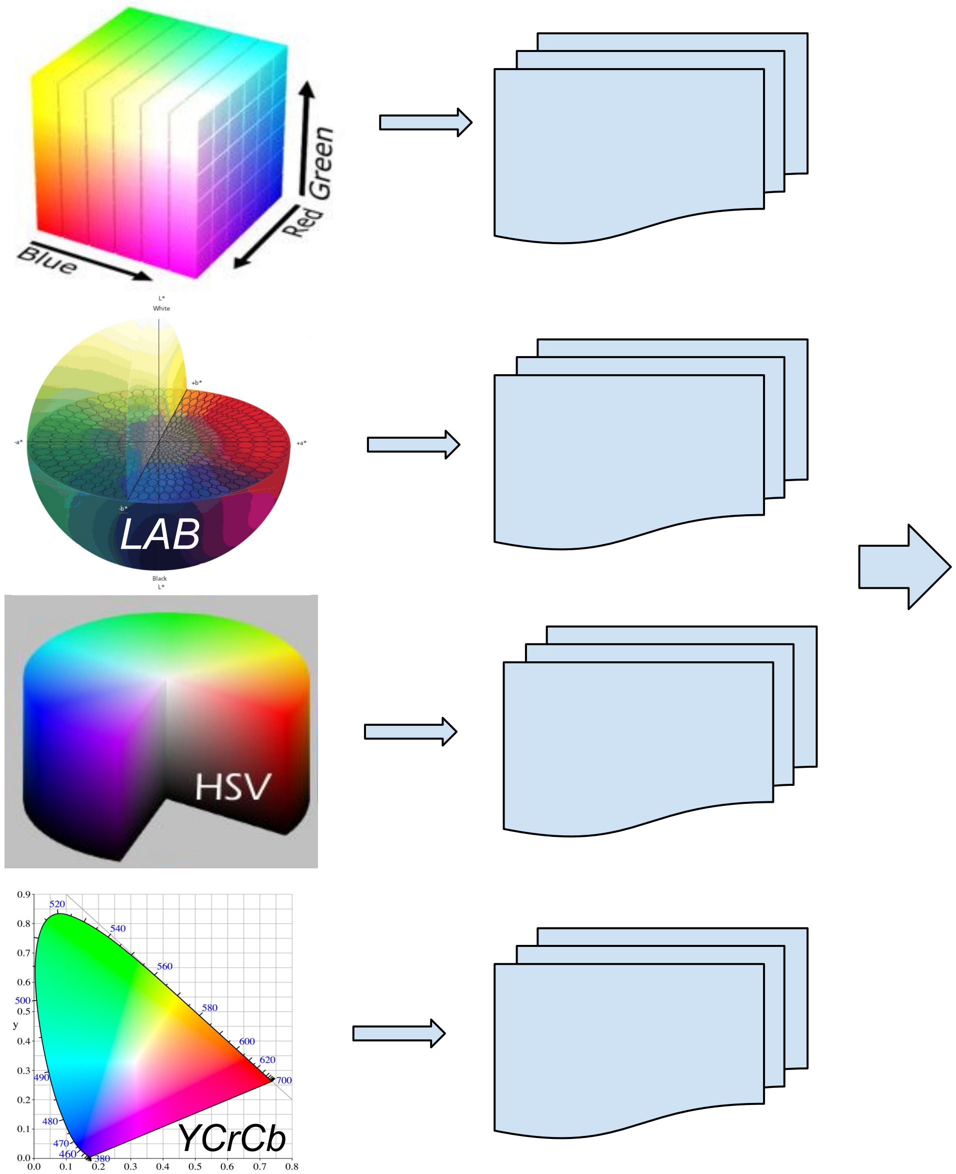
Segmentation algorithm fails for the three class problem







Features Extraction: Color



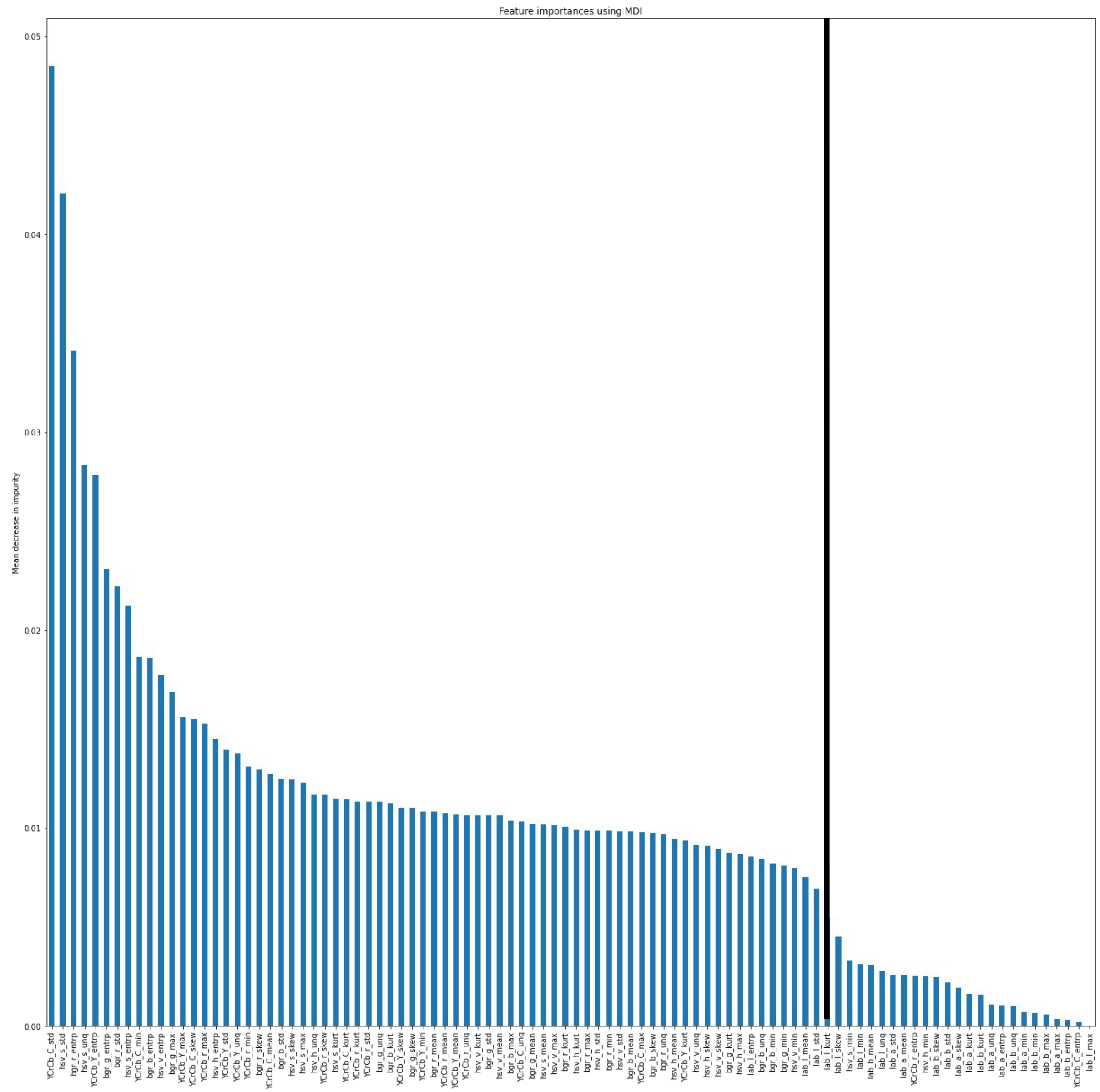
Channel-wise color features

1. Mean
2. Variance
3. Skewness
4. Kurtosis
5. Max
6. Min
7. Entropy
8. Number of unique values

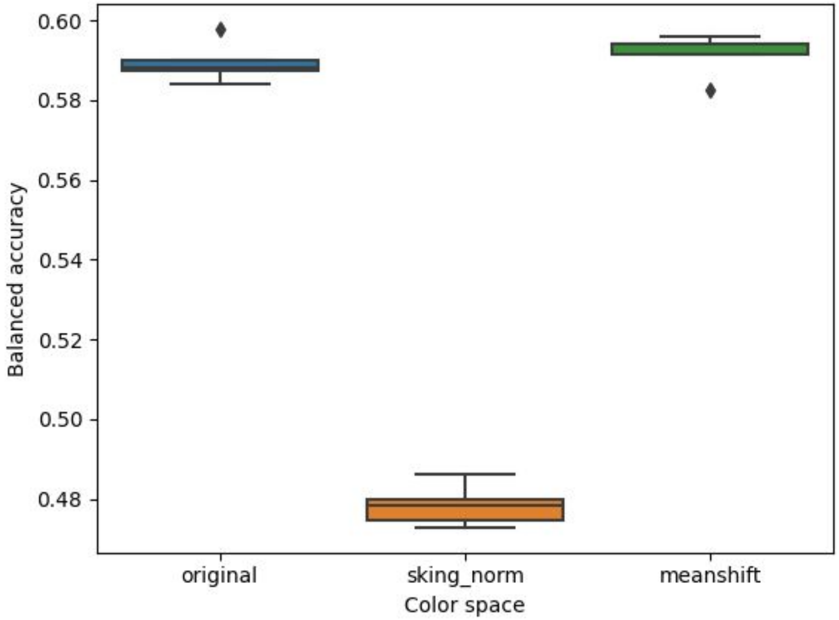
Universitat de Girona

Features Extraction: Color

- Lab* colorspace features were the weakest (removal of these features led to the improvement of the weighted f1 from 0.7881 to 0.7974) and decreased number of features from 96 to 72

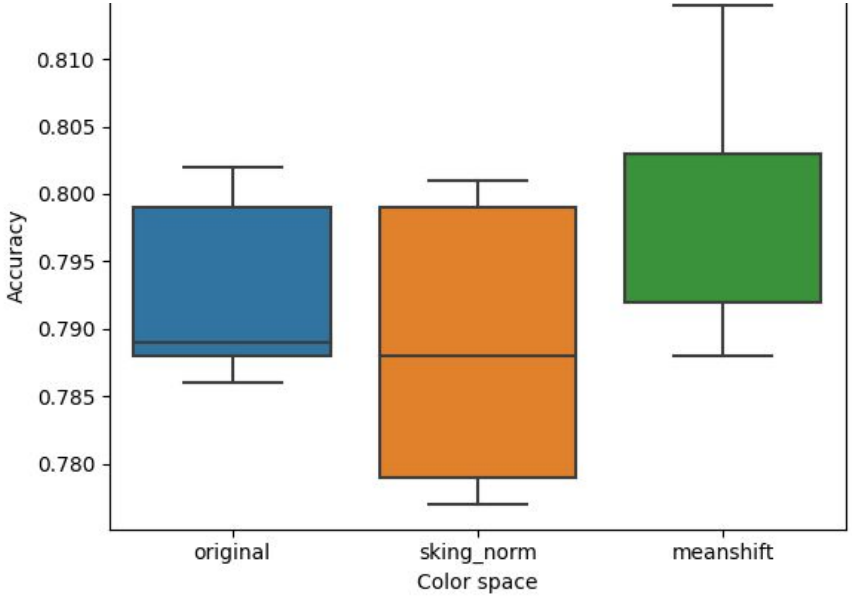


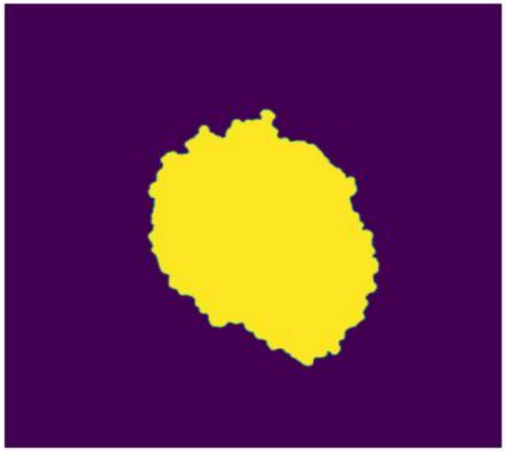
Preprocessing: Color Normalization



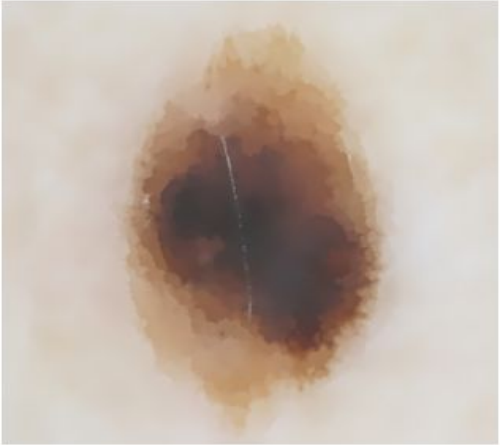
Original Image

Skin Color Normalization





Segmentation



Meanshift

Features Extraction: Texture

GLCM with:

- Distances [2, 5, 7, 10, 15]
- Angles [0, 45, 90, 135]

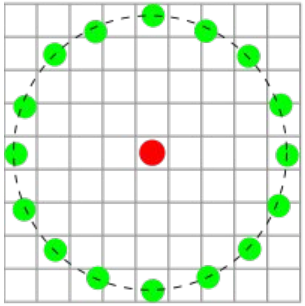
GLCM examples

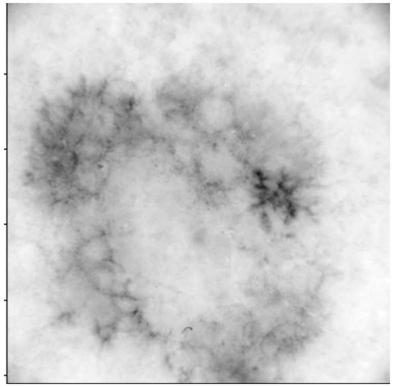
GLCM features

1. Contrast
2. Dissimilarity
3. Homogeneity
4. Energy
5. Correlation
6. Angular Second Moment (ASM)

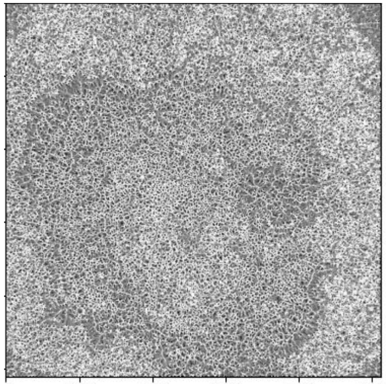
LBP histograms

9 different radius and
number of points combinations

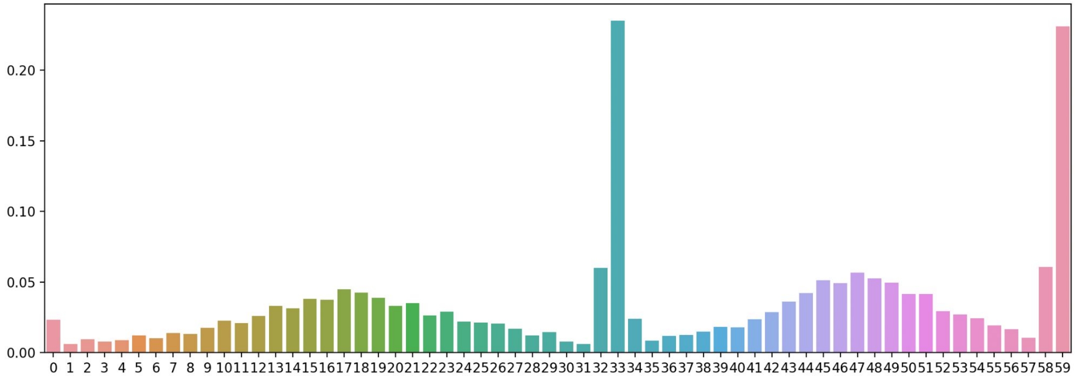




Gray scale image



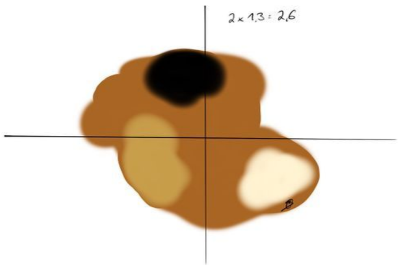
LBP image

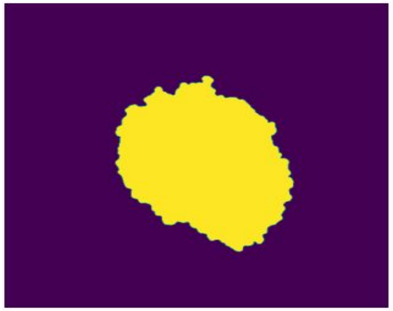


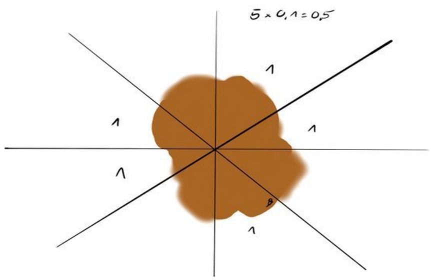
Universitat de Girona

Feature Extraction: Shape

Asymmetry and border features

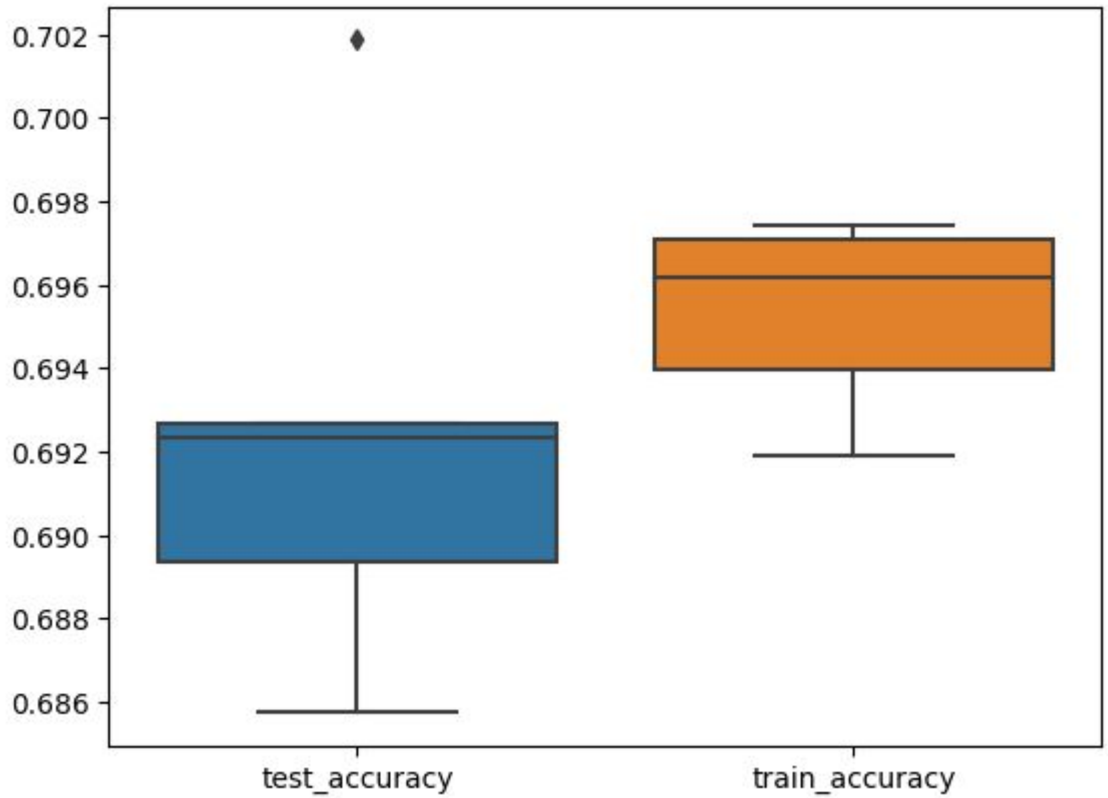






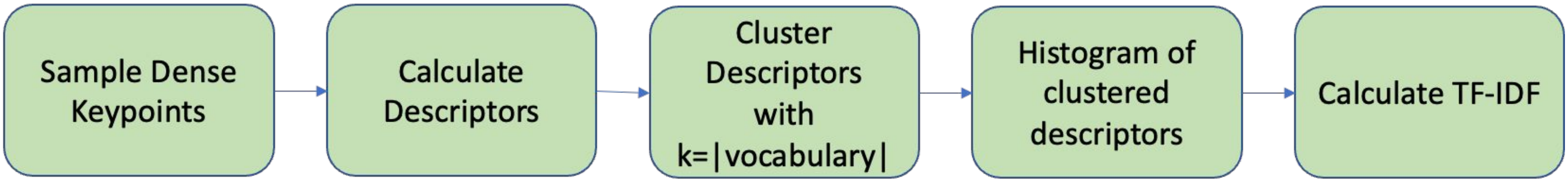
1. Number of lesions in the mask
2. Mean and std of their areas
3. Area
4. Perimeter
5. Circularity
6. Eccentricity
7. Aspect ration
8. Compactness index
9. 7 hu moments

Five-fold CV on full train set of challenge 1 results on only shape features



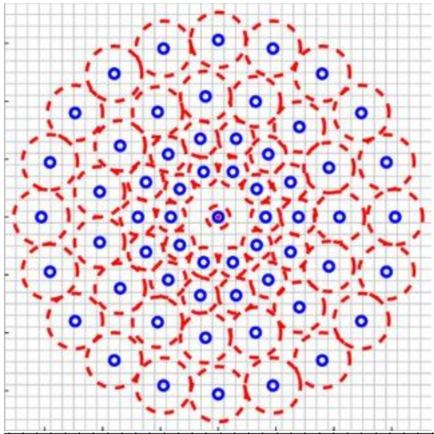
Universitat de Girona

Feature Extraction: BoW

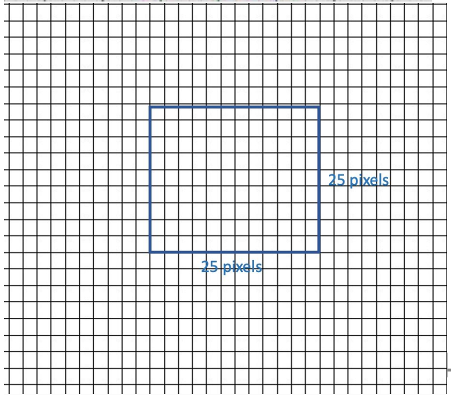


Descriptors experimented with:

1. Brisk: constructs the feature descriptor of the local image through the gray scale relationship of random point pairs in the neighborhood



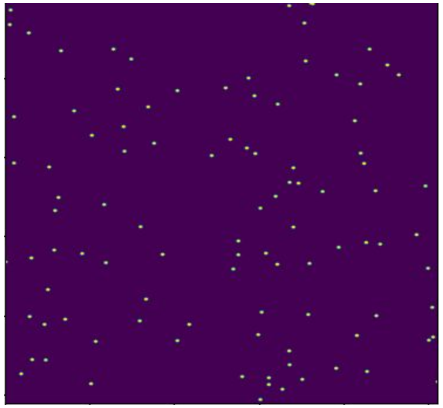
2. Color, GLCM, LBP: Calculate the features within patch size of 25 centred at the keypoint



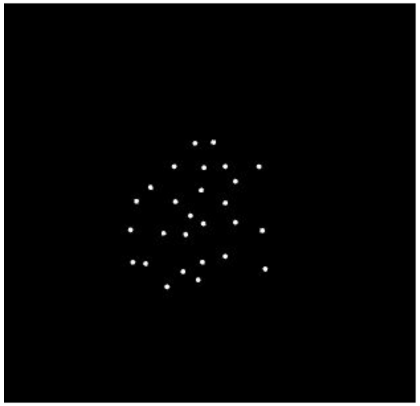
Universitat de Girona

Feature Extraction: BoW

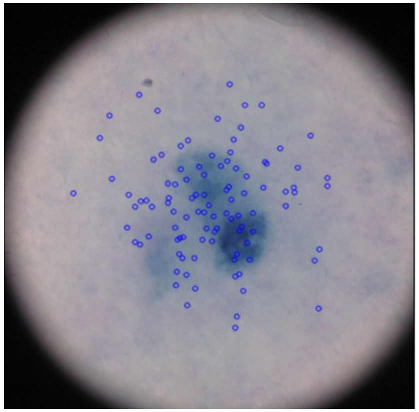
Keypoint Sampling Strategy	Accuracy for Texture Descriptors (challenge 2)	Accuracy for Color Descriptors (challenge 2)	Comments
1. Random within segmentation mask	0.5818	0.6323	Segmentation not good enough for challenge 2
2. Random within centered radius as mask (radius 100)	0.5717	0.6606	Better for color features
3. Gaussian sampled at the centre of the image	0.5959	0.62424	Better for texture features



1



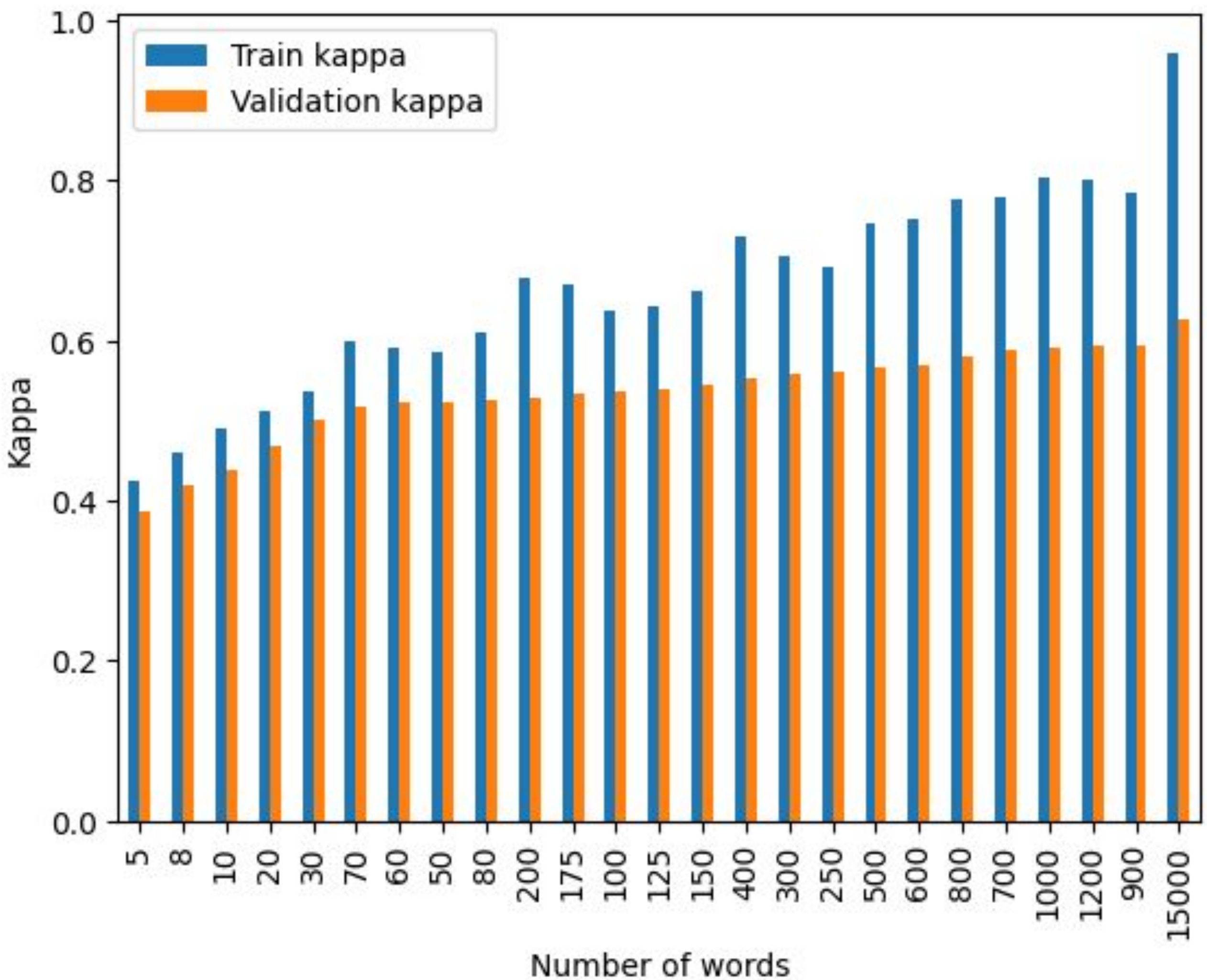
2



3

Feature Extraction: BoW

Vocabulary size experiment: 100 words are enough



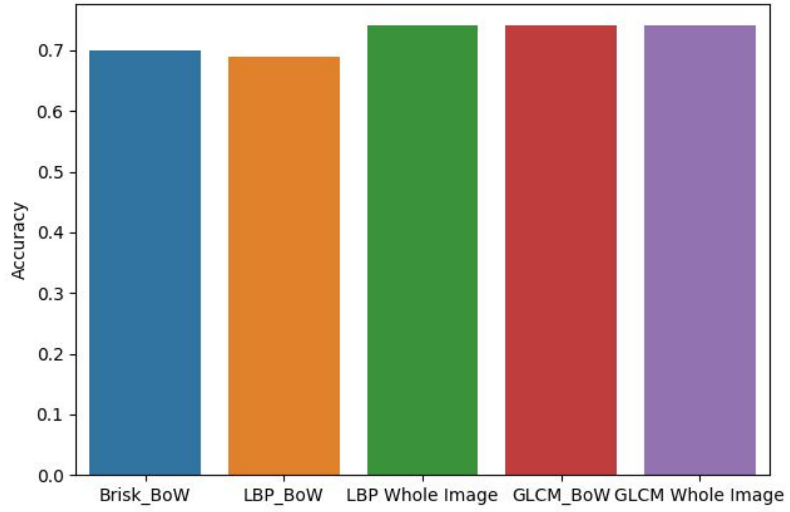
Feature Extraction: BoW

Binary Problem

- BoW not better than whole image features

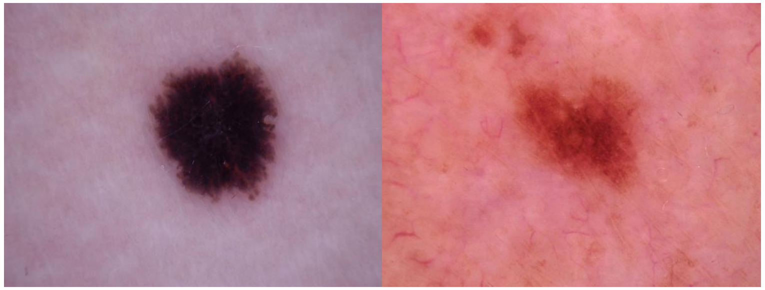
3 Class Problem

- BoW Improved the validation accuracy by ~0.5

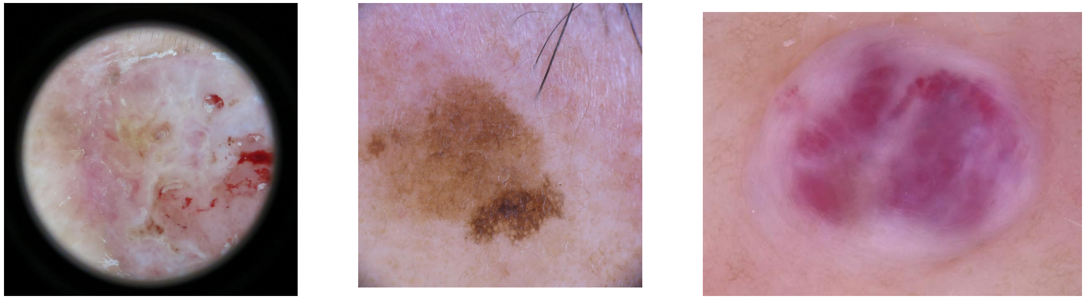


Universitat de Girona

Challenge 1: Overview and Features



Nevus Images



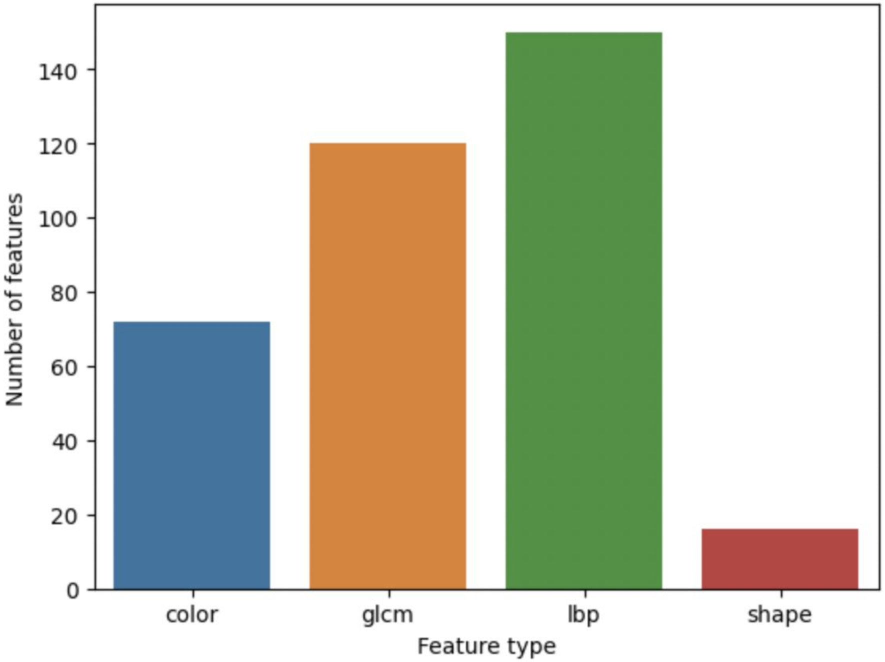
Others Images

Overview: Binary classification problem; balanced huge dataset

Total: 358 features of color, texture and shape

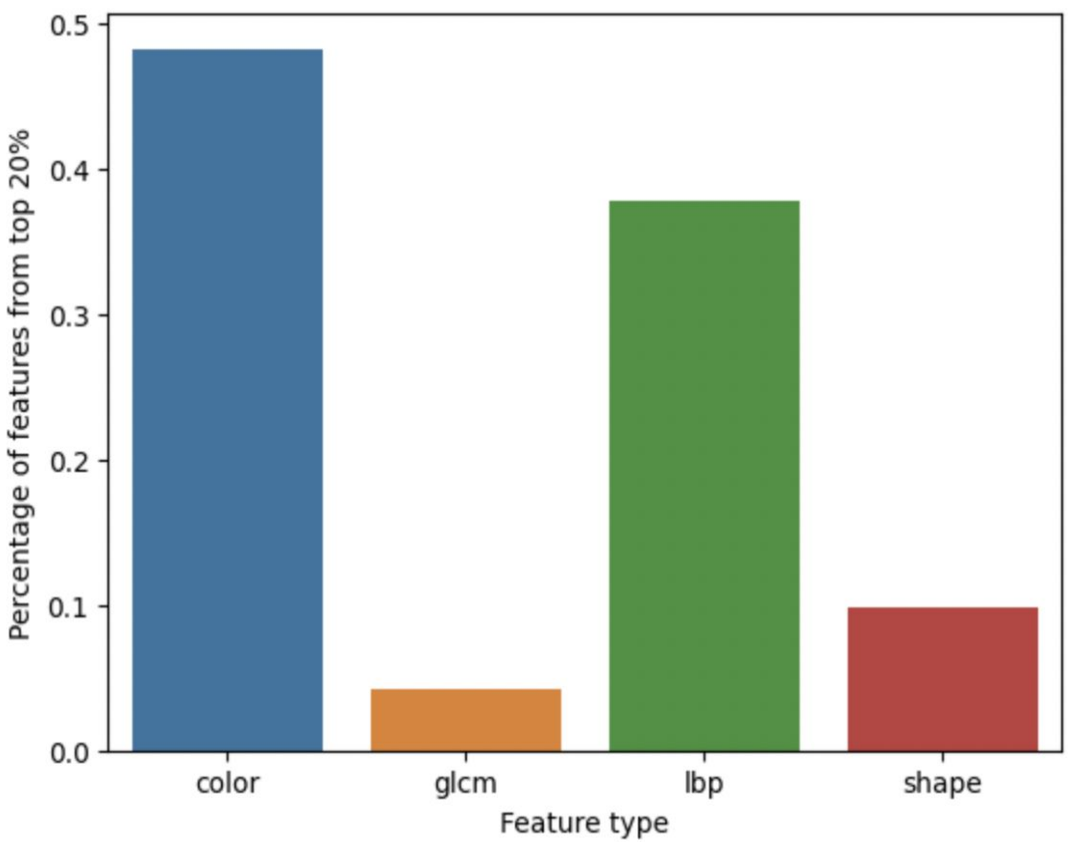
No BoW features didn't bring significant improvement

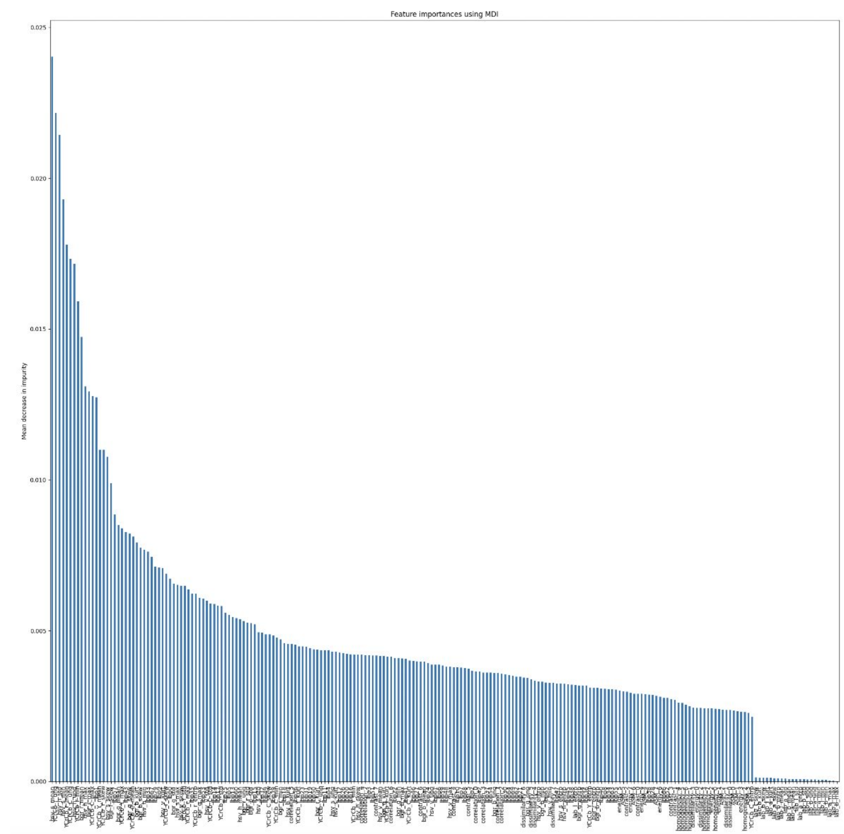
Explored: reducing feature size to tackle the curse of dimensionality



Challenge 1: RF Feature Selection

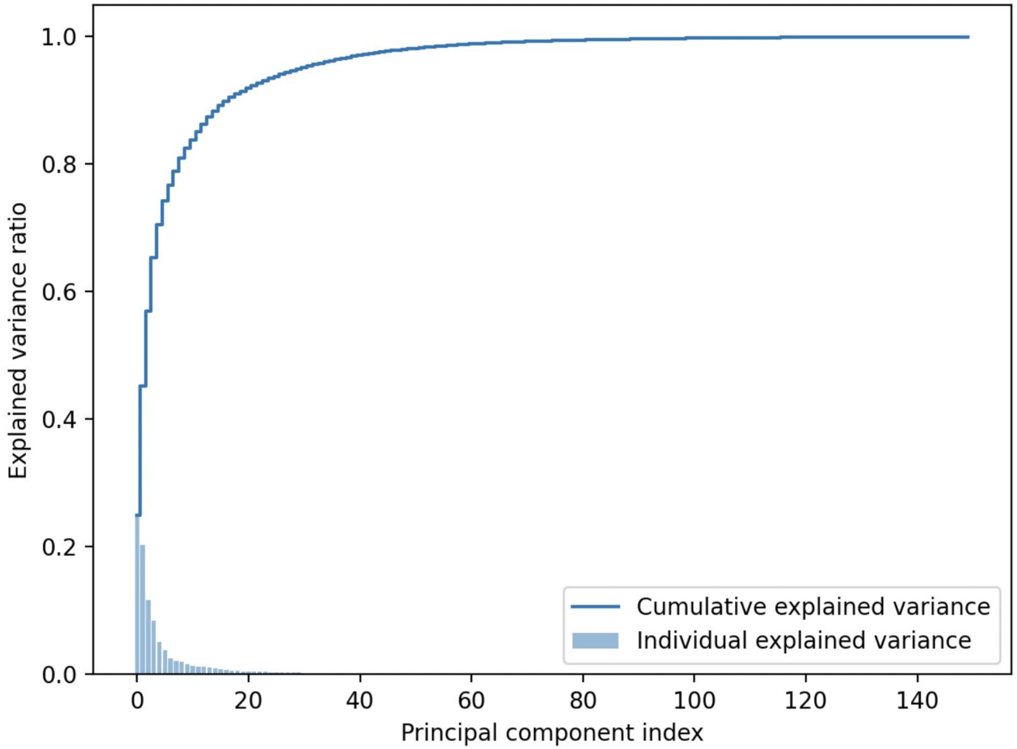
- Lab color space features were removed.
- Selecting k-best features didn't improve the validation accuracy

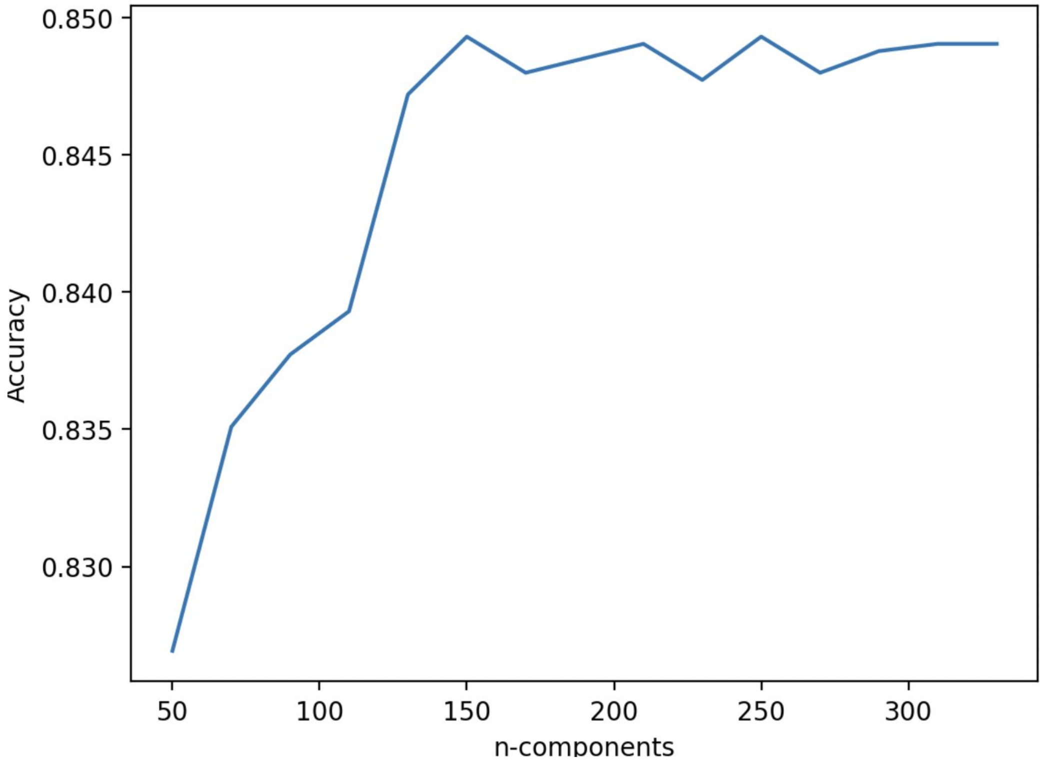




Challenge 1: PCA Dimensionality Reduction

150 principal components chosen as final set of features from 358 features

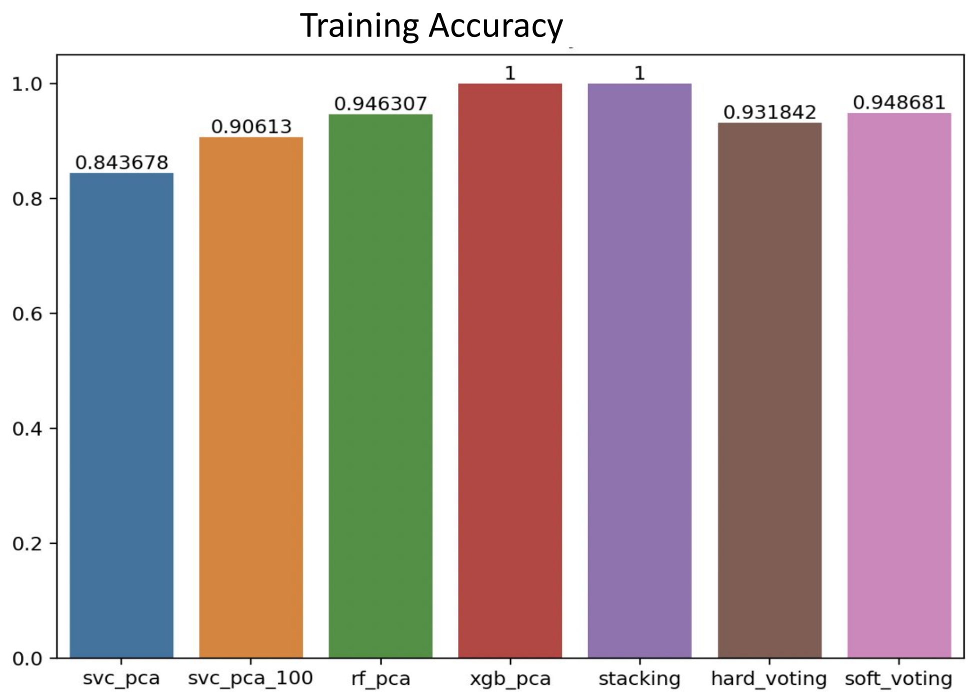


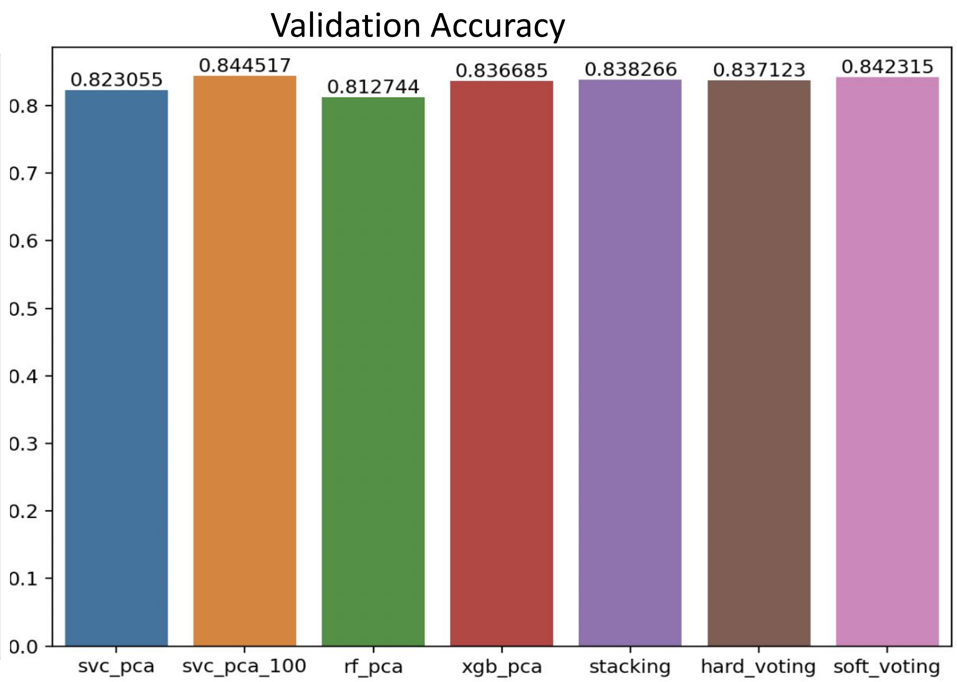


Universitat de Girona

Challenge 1

- All extracted features reduced with PCA (150 components) were used
- Soft-voted Ensemble of tuned SVM and XGBoost classifiers was used as the best trade-off between training (less overfitting) and validation(generalization) accuracy.





Universitat de Girona

Challenge 2: Overview and Features

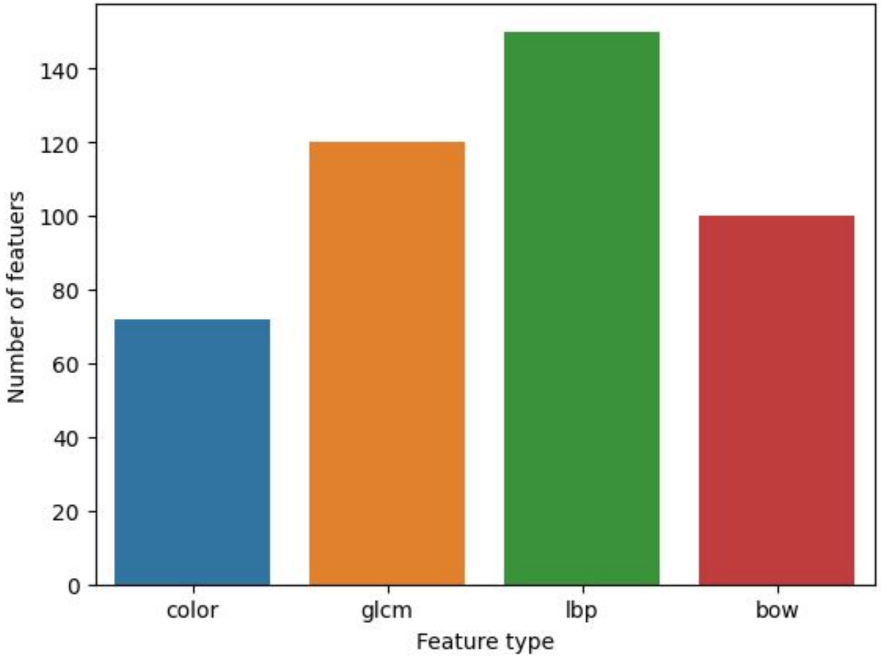
Basal Cell | Squamous Cell | Melanoma

Challenges: multi class, less data, highly imbalance data set.

Total: 442 features of color (both global and BoW tf-idf) and texture

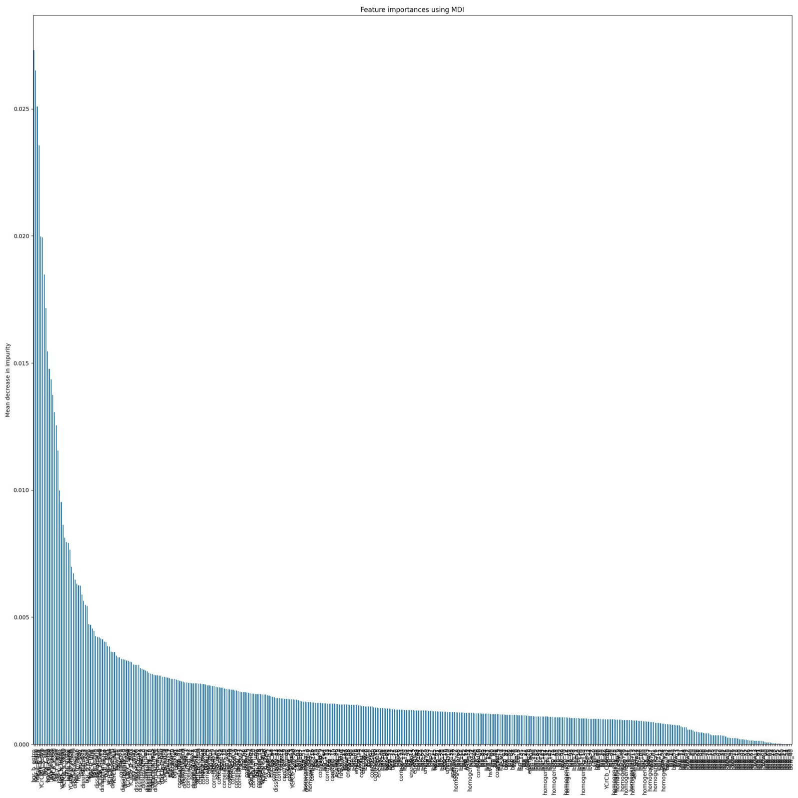
No shape features since the segmentation results were poor

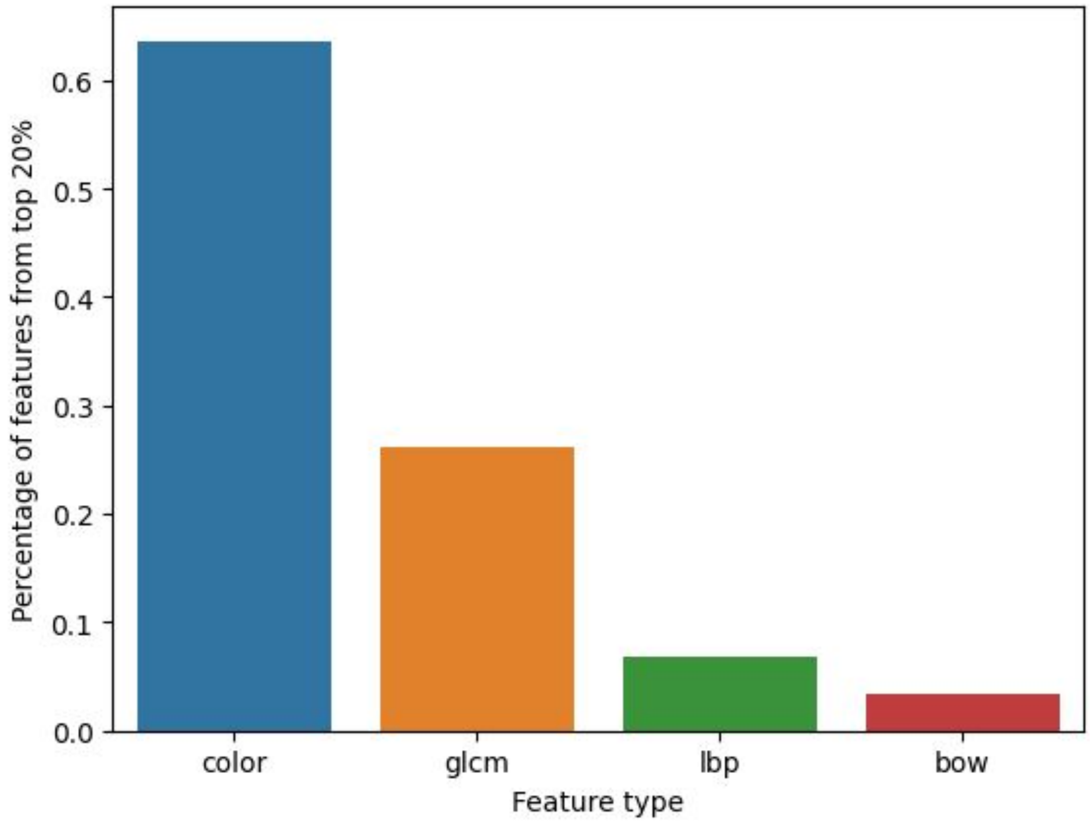
Explored: reducing feature size to tackle the curse of dimensionality and techniques to sive the imbalance



Universitat de Girona

Challenge 2: RF Feature Selection





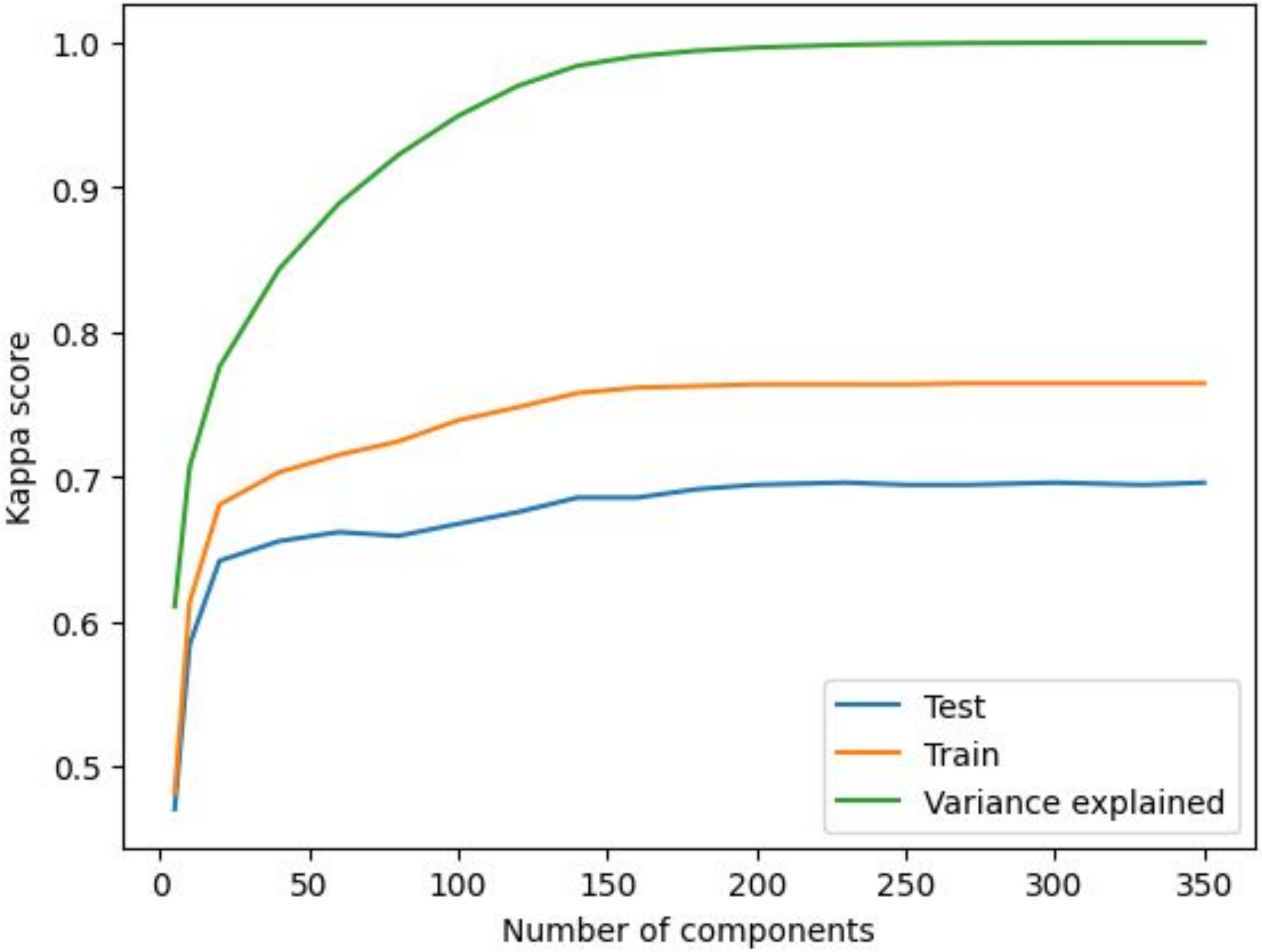
Analysis of feature importances of top 20% RandomForest features shows that global color and glcm texture features were the most prominent ones with BoW and LBP features still having an important contribution.

A further investigation of a features set composed of these 88 top 20% features was done (referred to as rf_fs) .

Challenge 2: PCA Dimensionality reduction

Validation set kappa score for the different number of components in PCA decomposition of all 442 features.

A further investigation of a features set composed of these 150 PCA features was done (referred to as pca) .

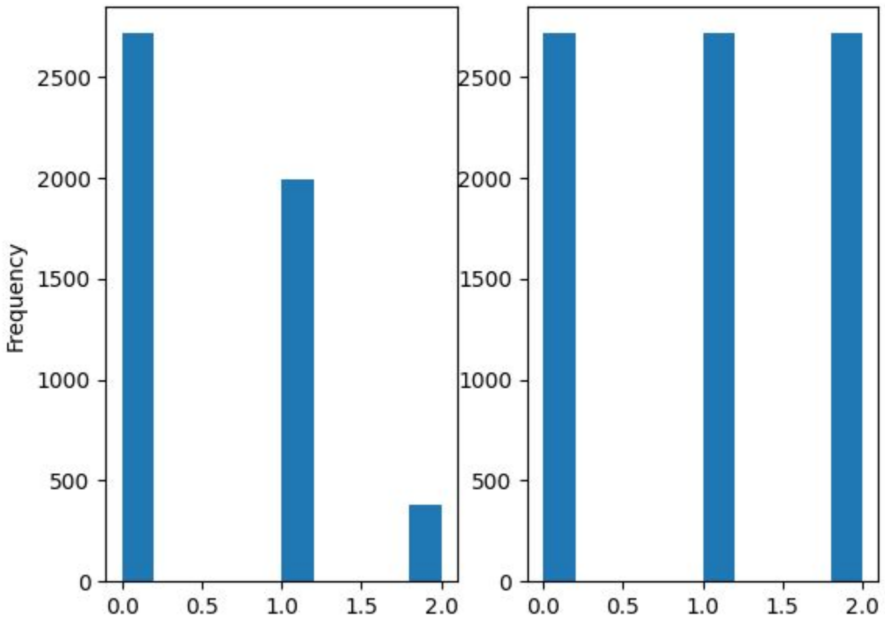


Challenge 2: Imbalance Problem

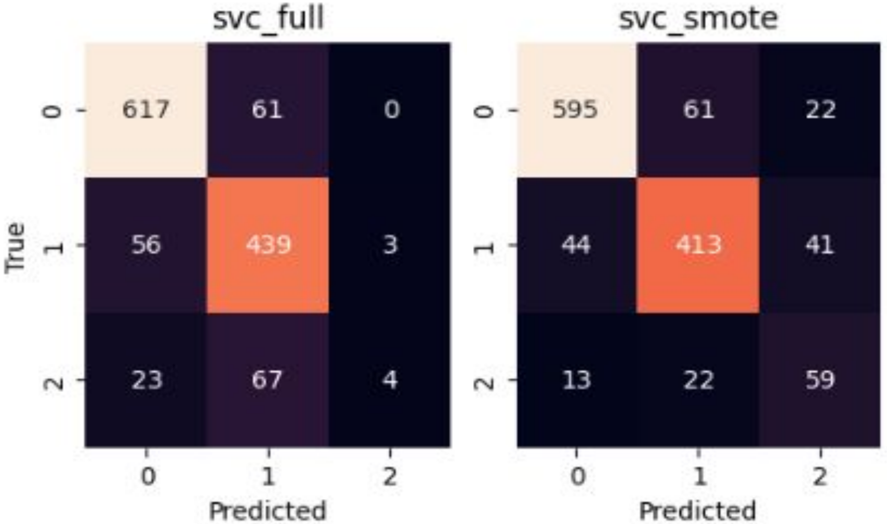
Data-wise we have tried:

- oversampling
- undersampling
- Synthetic Minority Oversampling Technique (the only one to show any improvement)

Model-wise: use of <u>balanced class weights</u> in all of the classifiers we were testing

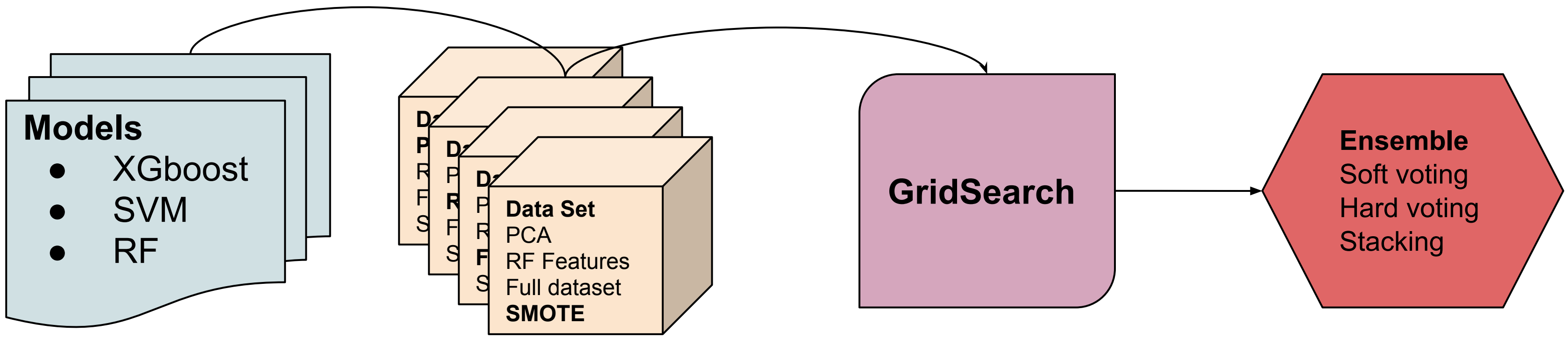


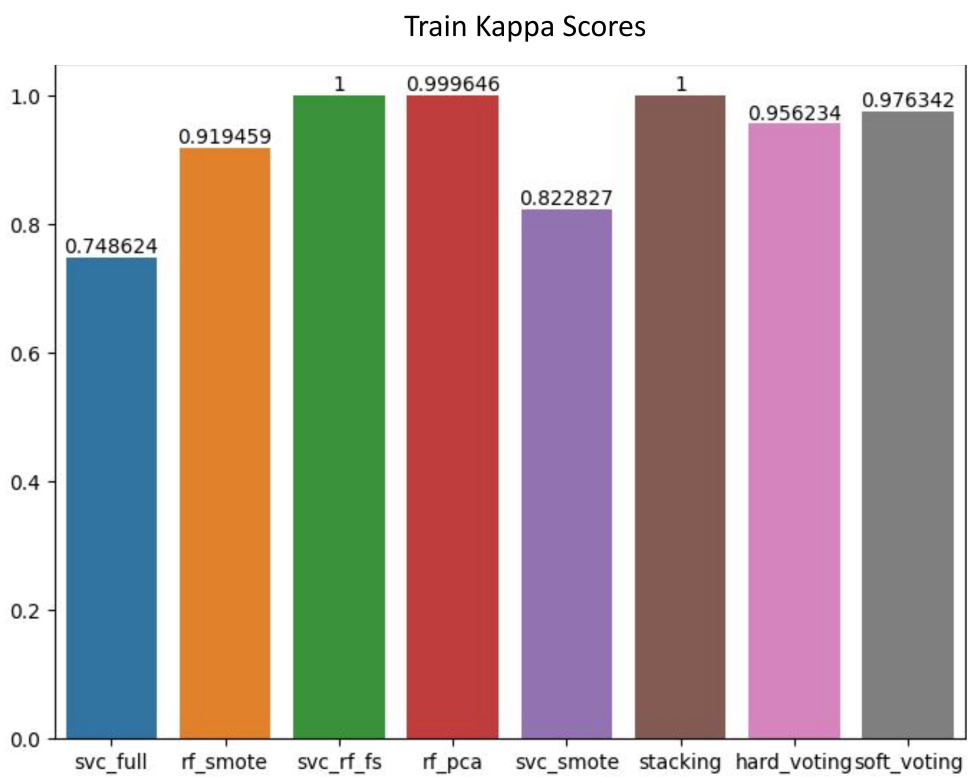
Train set class distributions before and after SMOTE

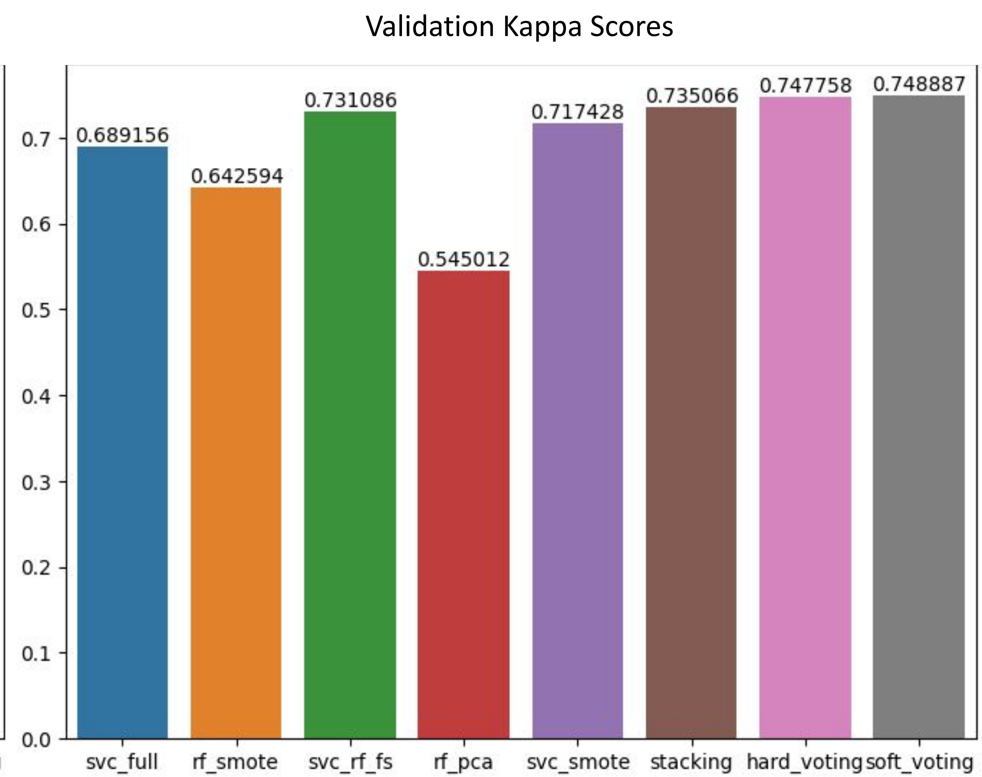


Validation set Confusion Matrices

Challenge 2: Ensembling





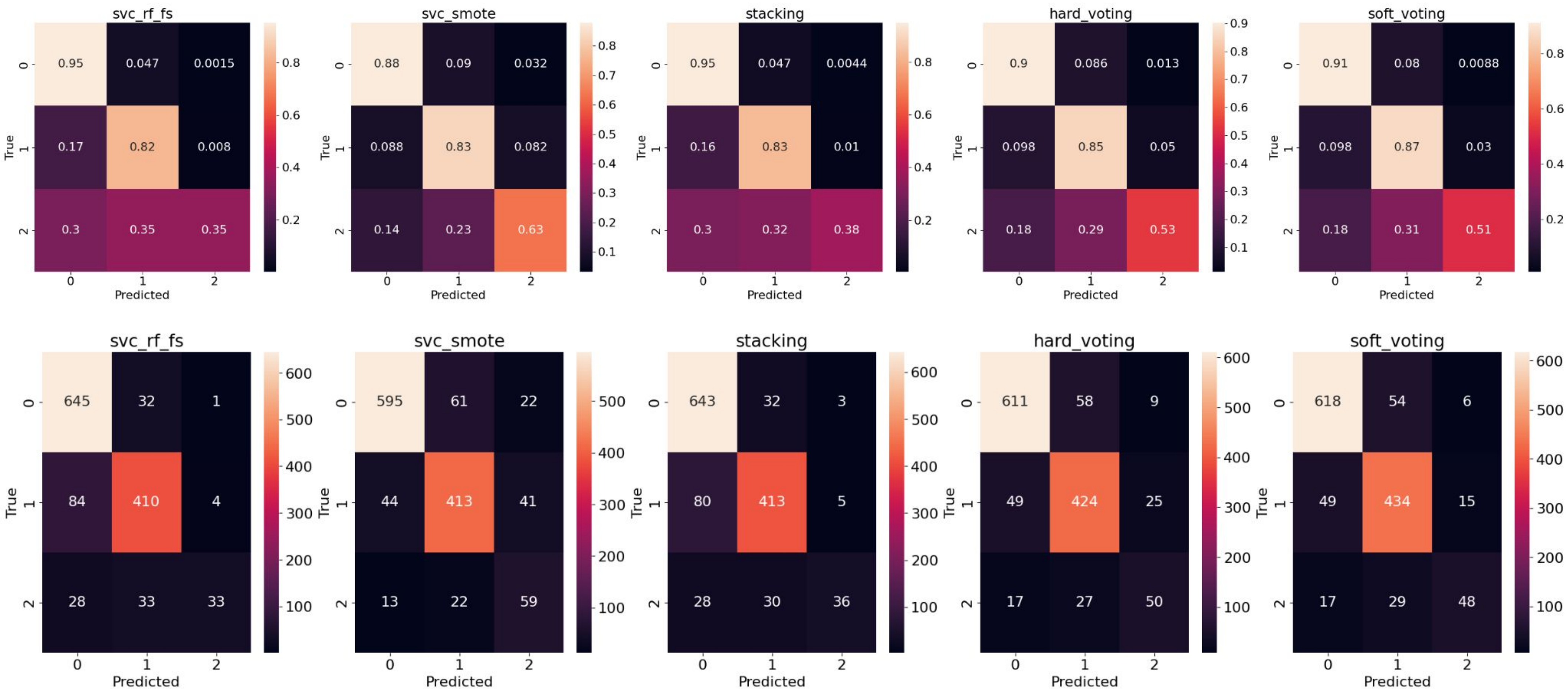


MAIA

Challenge 2: Final model

Taking a closer look at the confusion matrices of the top 5 model/dataset combinations from the previous slide we can notice that **SVM on SMOTE data achieved the best results**. It has the <u>smallest overfitting</u> while maintaining <u>the best proportion between 3 classes.</u>

Therefore, we have selected it as our final model for the challenge 2 with the validation **kappa of 71.**



Universitat de Girona

Conclusions

- Color features are the most discriminative for both problems
- Segmentation of lesions can lead to better results but is quite challenging, especially for malignant lesions
- BoW was able to improve performance on 3 class problem, due to increased importance of the small variations in color information between lesion types
    - However for 2 class problems global color features were more effective
- Adding additional features (like texture, shape or BoW) improved the results however also led to increased overfitting
- Data imbalance for 3 class problem was better solved with balance weights SVM on SMOTE data, however the minority class still was considerably underdetected
    - needs more distinctive features