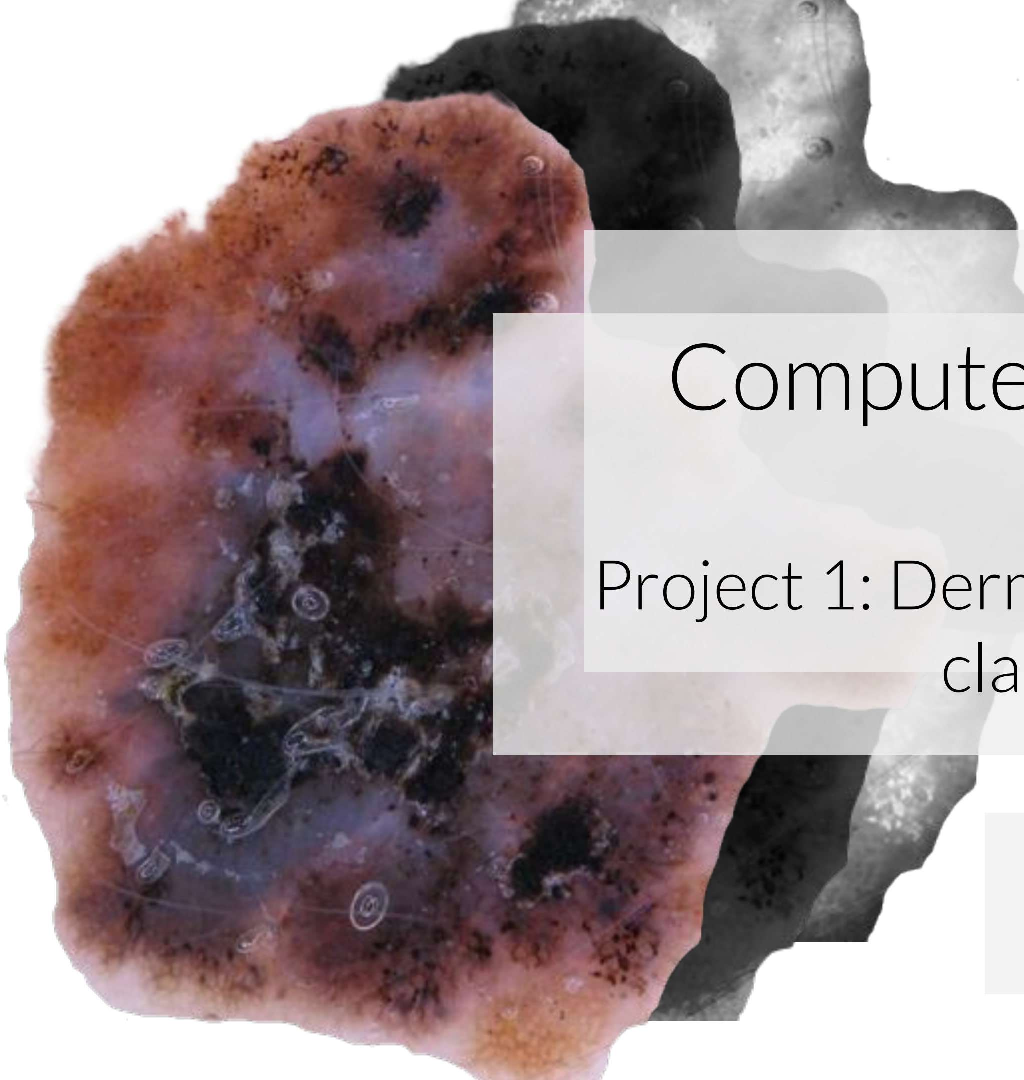


MAIA

Universitat de Girona

Computed Aided Diagnosis

Project 1: Dermoscopic diagnosis with the classical approach

Alejandro Cortina Uribe
Cansu Yalcin

Contents

Dataset description

- Challenge 1: Binary

# images	Nevus	Others							
	Total	Total	mel	bcc	bkl	ack	scc	vac	def
Train	7725	7470	2713	1993	1574	520	376	151	143
Val	1931	1865	678	498	393	130	94	37	35

Balanced dataset

Different image sizes

- Challenge 2: Three class

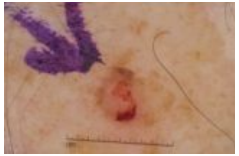
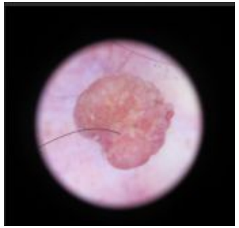
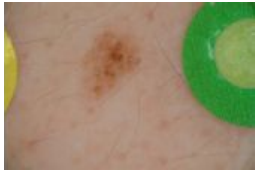
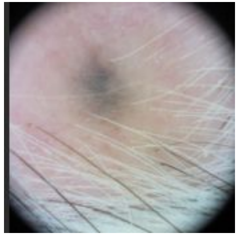
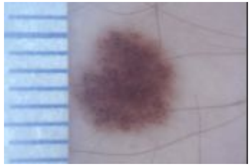
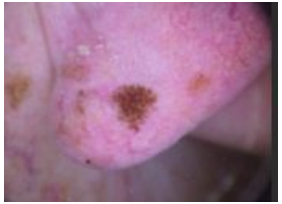
# images	bcc	mel	scc
Train	1993	2713	376
Val	498	678	94

Unbalanced dataset

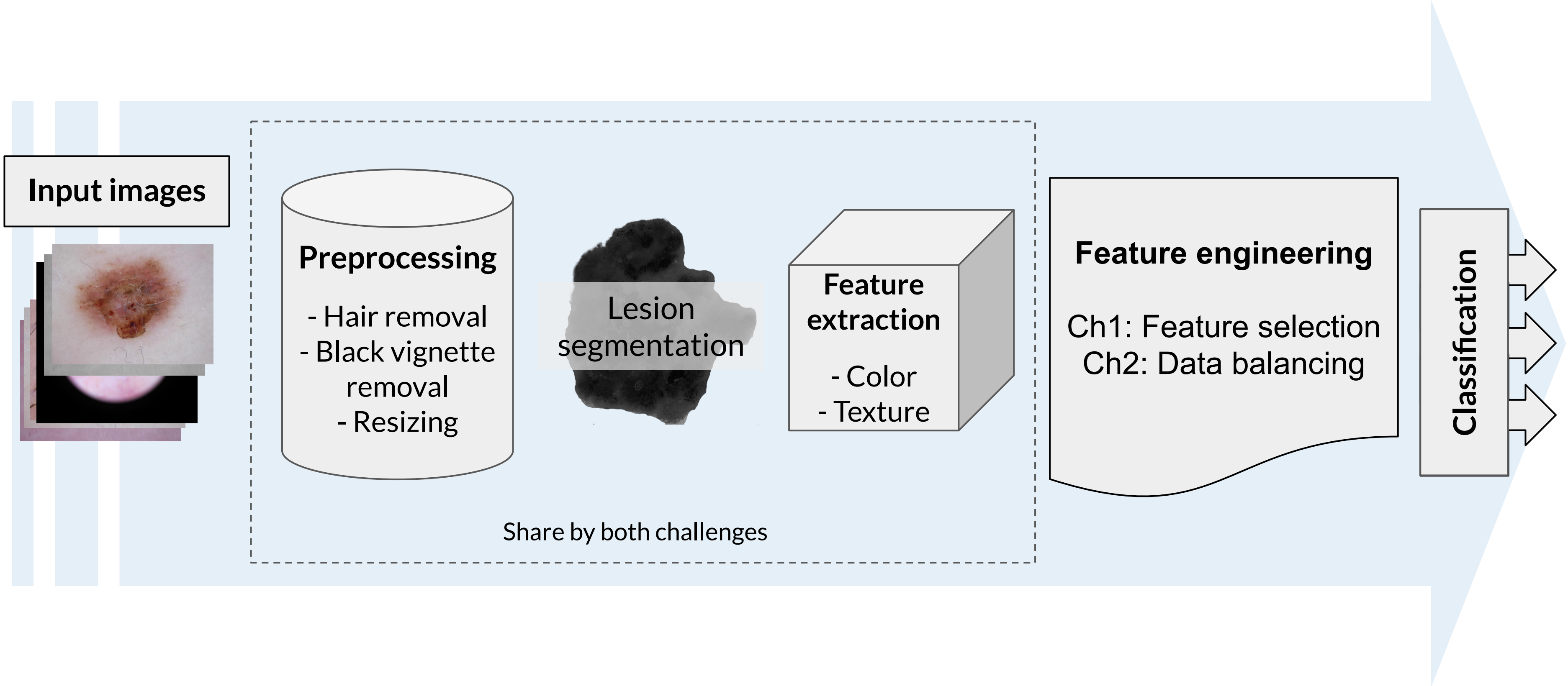
Different image sizes

Image characteristics:

- Artifacts:
  - Dark and light hairs
  - Marker annotations
  - Rulers
  - Stickers
- Black circular vignette (border)
- Variate non-standard lesion framing.

Pipeline



Input images
Preprocessing
- Hair removal
- Black vignette removal
- Resizing
Lesion segmentation
Feature extraction
- Color
- Texture
Share by both challenges
Feature engineering
Ch1: Feature selection
Ch2: Data balancing
Classification

Preprocessing: Black vignette removal

Algorithm[1]:

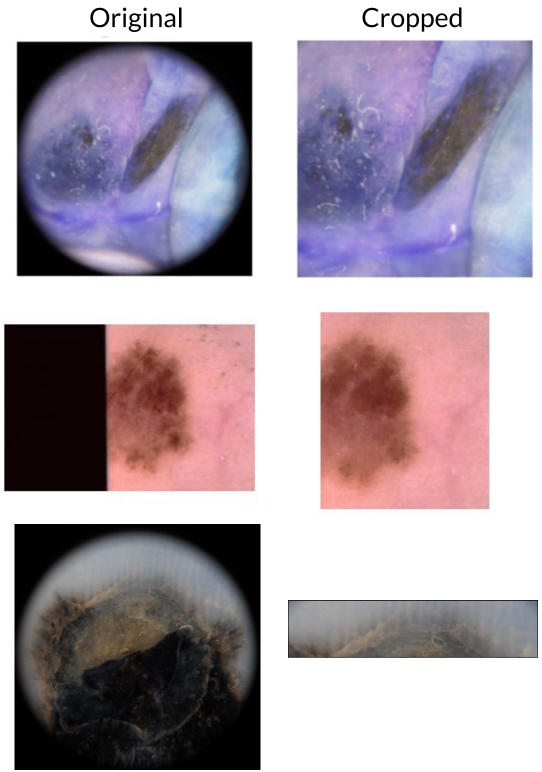
1. Obtain the histogram of image diagonals
2. Obtain histograms crossing points with a selected threshold
3. Select the most restrictive cropping coordinates

Problems:

- Non-circular vignette (single horizontal or vertical black border)
- Off-center nearly-black lesion

Solutions:

- Remove non-circular vignette
- Obtain very restrictive crop



Preprocessing: Hair removal

Algorithm[2]:

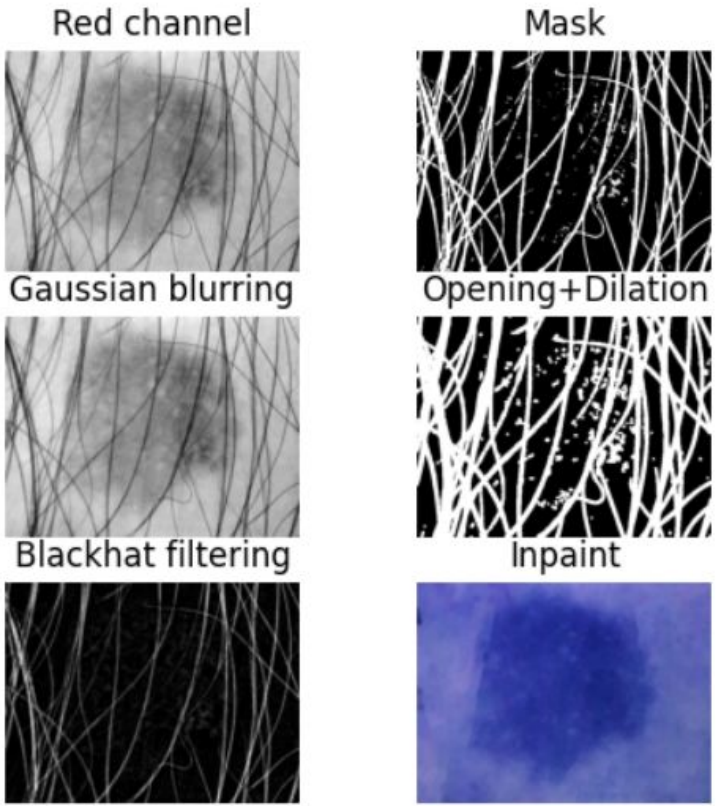
1. Gaussian blurring followed by black hat filter in the red channel
2. Binary thresholding
3. Opening + Dilation morphological operations
4. Inpainting

Problems:

- White hairs (not solved)
- Losing lesion information
- Computational time

Solutions:

- Setting conservative parameters to avoid influence of problems

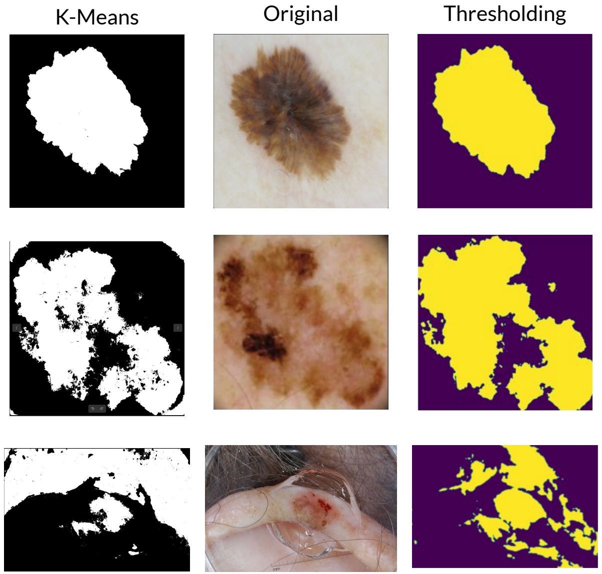


Lesion segmentation

Goal: Obtain a good-enough lesion segmentation to later extract **local** features and lesions' **shape** features (i.e. according to the ABCD of visual classification of lesions[3]).

K-Means[4]

1. Implement K-Means algorithm with:
- K = 2
- Attempts = 10
- Iterations = 100
- Epsilon = 0.2



Global thresholding[5]

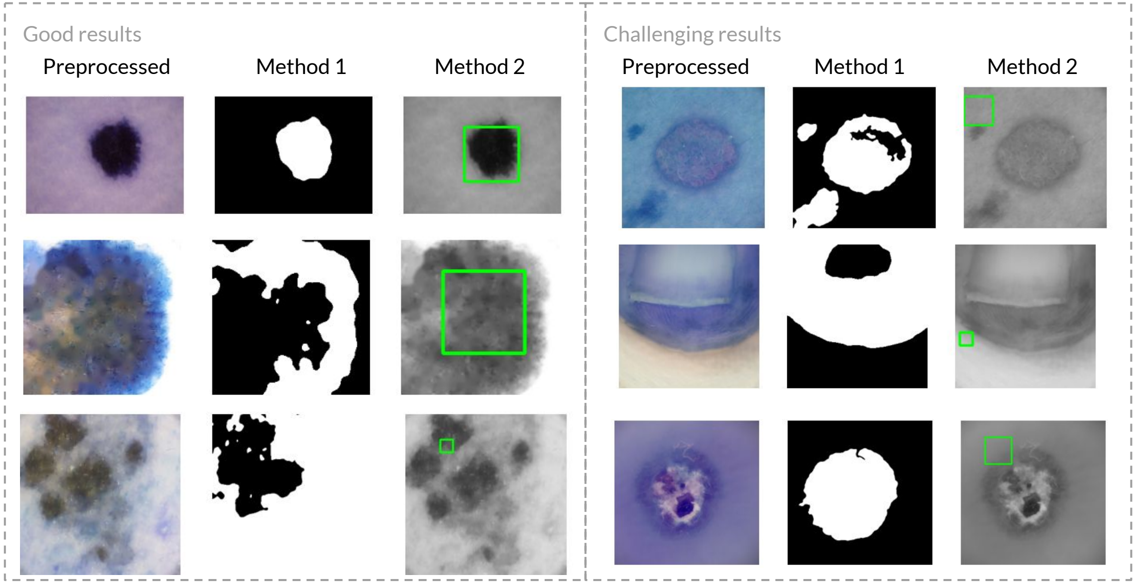
1. Get normalized R channel from RGB image
2. Blur with Gaussian kernel
3. Threshold with Otsu
4. Fill holes

*Shown k-means results were obtained without preprocessing (hair and black frame removal)

Lesion segmentation

Problem: Due to lesion type characteristics, such as irregular blurry edges, non-uniform color/texture inside the lesion, non-circumscribed margin, other non-lesion elements (anatomical structures), and artifacts, we focus only to **obtain a ROI** for local features extraction.

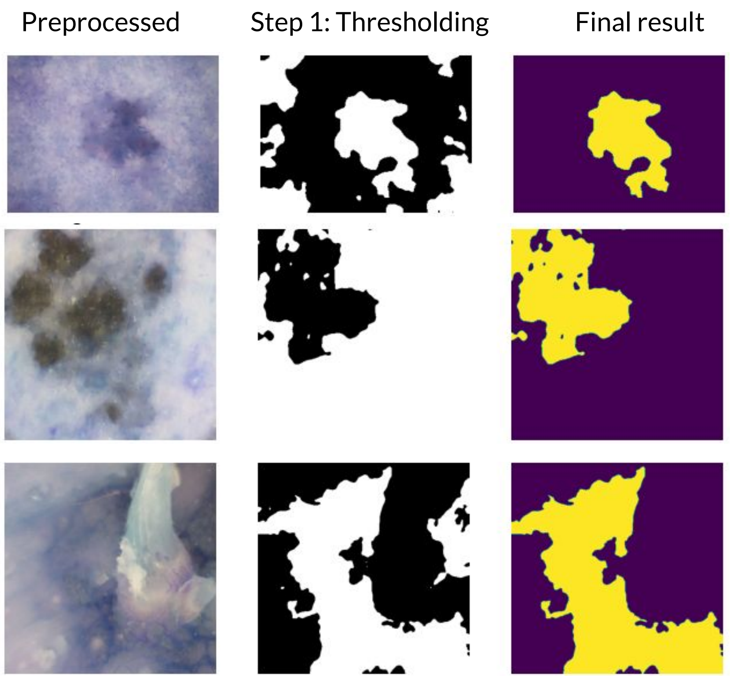
- Method 1:
  - Lesion segmentation by global thresholding (previous slide).
- Method 2:
  - Circular Hough transform
  - Obtain centroid of largest detected circles



Lesion segmentation - ROI

Solution: Improve best method.

1. Lesion segmentation by global thresholding
2. Get connected components (4-connectivity)
3. Obtain largest component that it is not background
4. If algorithm’s conditions are not met, return binary mask from step 1.



Feature Extraction - Color Statistics[4, 6]

Goal: Obtain first order statistics from different color spaces for each channel.

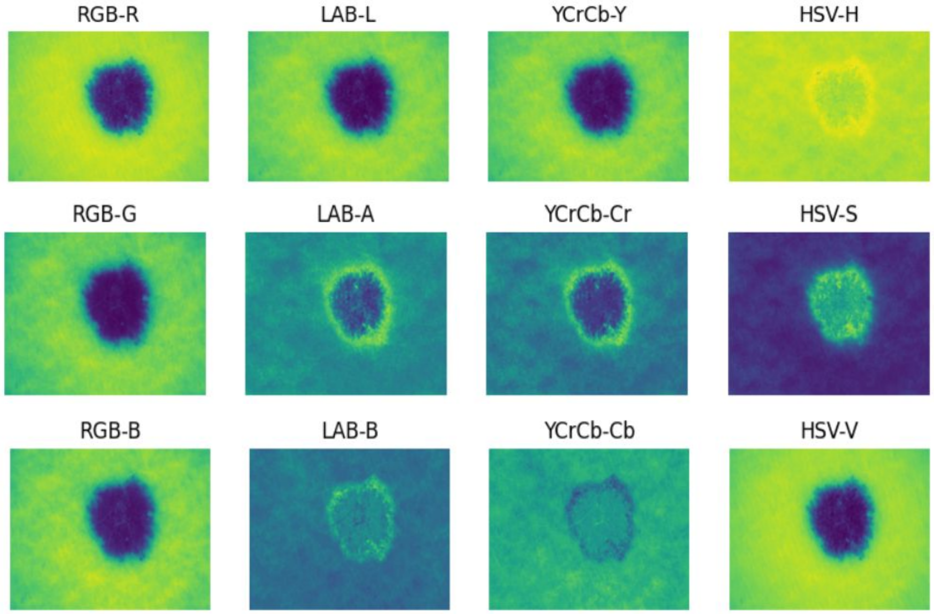
Color spaces:

- RGB, LAB, HSV and YCrCb.

First order statistics:

- Mean, Standard Deviation, Skewness, Entropy, Kurtosis.

Overall, 60 global (full-image) and 60 local features (using segmentation) are extracted.



Feature Extraction - Local binary pattern (LBP)[7]

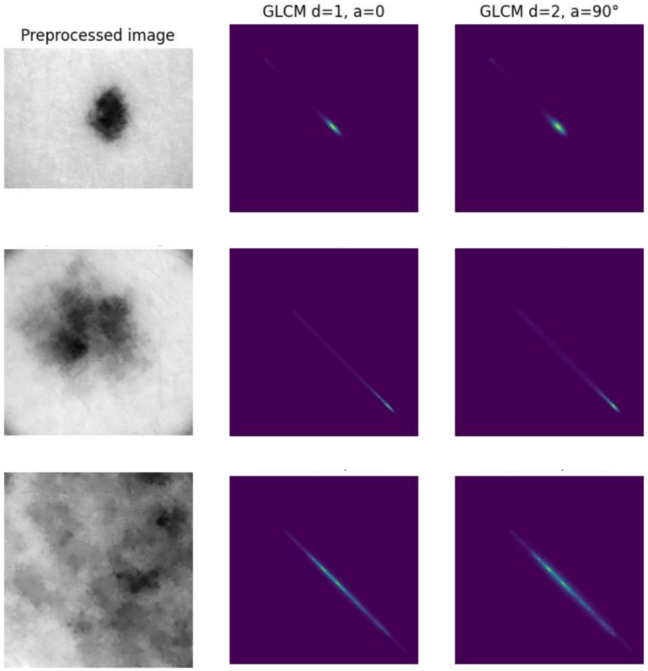
Goal: Extract texture features (histogram of patterns) that are gray scale and rotation invariant.

- Parameters
  - Type: uniform
  - Radius: 1 and 3
  - Points: 8 and 24, respectively
- Get the normalized histogram of the obtained lbp's.
- We tested extracting features for each channel in different color spaces, but information was redundant.
- Overall 20 features are extracted, from grayscale (single-channel) image.

Feature Extraction - Gray Level Co-occurrence Matrix (GLCM)

Goal: Extract statistics from different GLCM's of the gray scale image [3, 8]

- Parameters
  - Angles: 0, $\pi/4$, $\pi/2$, $3\pi/4$
  - Distance: 1 and 2, for each angle
- Once the GLCM's are computed, obtain the statistics: contrast, dissimilarity, homogeneity, ASM, energy, and correlation.
- Overall 48 features are extracted, from grayscale (single-channel) image.



Feature Extraction - Gabor Filters[9]

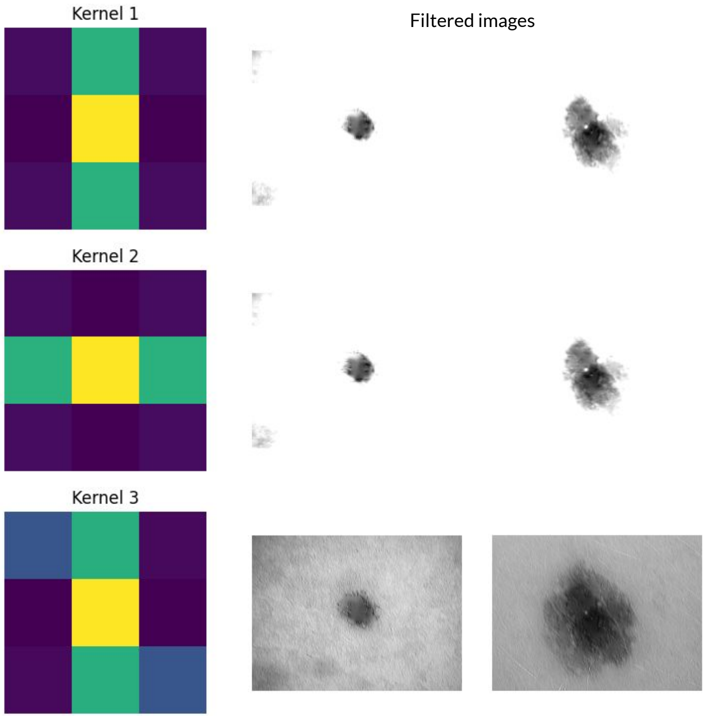
Goal: Calculate the global image statistics after Gabor filter application.

- Gabor kernels are tailored to capture small and circular features in the images.

First order statistics:

- Mean, Standard Deviation, Skewness, Entropy, Kurtosis, Minimum

Overall, 15 features are extracted.



Classification - Classifiers

We performed several classification techniques. In order to find the best parameters, grid search is applied.

- Logistic Regression: Parametric, supervised learning classifier
- K-nearest neighbors: Non-parametric, supervised learning classifier
- Random Forest: Ensemble learning
- Adaboost: Ensemble learning
- Gradient Boosting: Boosting technique
- XGBoost: More regularized boosting technique
- Support Vector Machine Classifier: Supervised learning classifier
- Custom ensemble model: The combination of the best models.

Classification - Feature Selection

Two feature selection methods were tested from *sklearn.feature_selection* module.

Before feature selection the total number of global and local features was 203.

- VarianceThreshold: 151 features

  Feature selector that removes all low-variance features.

- SelectFromModel: 135 features

  Extracts best features of given dataset according to the importance of weights. LinearSVC is used as an estimator.

Classification - Experiments

- Evaluate the feature set
  - Global features
  - Global + Local

Classifier	Dataset evaluation	F1 Score		Accuracy	
		Train	Val	Train	Val
SVC	Global	0.9086	0.845	0.9086	0.8451
	Global + Local	**0.9359**	**0.8508**	**0.9359**	**0.8508**

SVC params: C = 10, kernel=rbf

- The table shows the results from SVC classifier with different feature sets. It proves that classifiers with Global + local features showed the better results.

- Evaluate the feature selection
  - VarianceThreshold
  - SelectFromModel
  - None

Classifier	Feature Selection	F1 Score		Accuracy	
		Train	Val	Train	Val
SVC	VarianceThreshold	0.9245	0.8416	0.9245	0.8416
	SelectFromModel	0.9257	0.8457	0.9257	0.8457
	None	**0.9359**	**0.8508**	**0.9359**	**0.8508**

- The table shows that without feature selection better results were obtained.

Custom Ensemble Model - Binary

- We have performed an ensemble learning model using the best binary classification models. It includes Random forest, XGBoost, and Support Vector Classifier (Table given in the Results part).

- Among with different experiments of combining the ensemble model, weighted voting with hard voting gave the best results.

- The weights were assigned according to the validation set performances of the classifiers.

Classifier	F1 Score		Accuracy	
	Train	Val	Train	Val
Hard Voting	0.993	0.8342	0.992	0.8342
Soft Voting	0.9875	0.8327	0.9874	0.8327
Weighted Voting (soft)	0.9866	0.8316	0.9865	0.8316
Weighted Voting (hard)	**0.993**	**0.8355**	**0.9929**	**0.8356**

Classification - Classifiers

We performed several classification techniques. In order to find the best parameters, grid search is applied.

- Support Vector Machine Classifier: Supervised learning classifier
- Random Forest: Ensemble learning
- Extremely Randomized Trees: Ensemble learning
- Adaboost: Ensemble learning
- Gradient Boosting: Boosting technique
- XGBoost: More regularized boosting technique
- Custom ensemble model: The combination of the best models.

Classification - Data balancing

To tackle the high imbalance of the dataset {bcc: 1993, mel: 2713, scc: 376}, we tested the following **balancing approaches**:

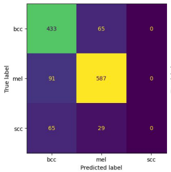
- Resampling with bootstrapping:
  - Down-sampling majority class: {bcc: 1993, mel: 1993, scc: 376}
  - Up-sampling minority class: {bcc: 1993, mel: 2713, scc: 1993}
  - Both: {bcc: 1993, mel: 1993, scc: 1993}
- Synthetic oversampling of minority class and undersampling without replacement of majority classes:
  - SMOTE: {bcc: 1000, mel: 1200, scc: 500}
  - ADASYN: {bcc: 1000, mel: 1200, scc: 685}
  - SVM-SMOTE (when main classifier is not a SVC)

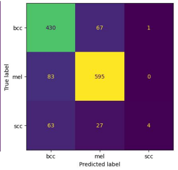
Classification - Support Vector Classifier

- Evaluate the feature set
    - Global features
    - Global + Local

Dataset evaluation (no balancing)	F1 Score (weighted)		Balanced accuracy score		Kappa	
	Train	Val	Train	Val	Train	Val
Global	0.7970	0.7736	0.5981	0.5784	0.6723	0.6299
Global + Local	0.8323	0.7847	0.6466	0.5945	0.7246	0.6434

SVC params. C = 10, kernel = rbf, gamma = 0.001





As SCC class gets better classification and overall the performance scores improve, we choose to continue the experiments with all features (global + local)

- Evaluate balancing techniques
    - No balancing
    - Resampling
    - Synthetic oversampling

Balancing	F1 Score (weighted)		Balanced accuracy score		Kappa	
	Train	Val	Train	Val	Train	Val
No	0.8323	0.7847	0.6466	0.5945	0.7246	0.6435
Up-sample minority	0.8684	**0.8024**	0.8657	**0.7421**	0.8007	0.6388
Down-sample majority	0.8234	0.7851	0.6563	0.5998	0.7197	**0.6535**
Both	0.8543	0.7855	0.8542	0.7310	0.7812	0.6062
SMOTE	0.8212	0.7848	0.7807	0.6704	0.7181	0.6116
ADASYN	0.8324	0.7846	0.8159	0.6922	0.7417	0.6116

SVC params. C = 10, kernel = rbf, gamma = 0.001

Custom Ensemble Model - Three class

Based on the overfitting degree of kappa score ($kappa_{train}$ - $kappa_{val}$) and the confusion matrices, we selected the following classifiers for the ensemble and carried out combinations of them with:

- Majority voting
- Soft voting
- Weighted soft voting

Estimator	Balancing	Notes
SVC	Down-sample majority class to medium class	The less overfitted model CM showing preference for **bcc** and **mel**
SVC	SMOTE	Less overfitted CM showing good balance for three classes
XGBoost	No	Overfitted CM showing preference for **bcc** and **mel**
RF	No	Very overfitted CM showing preference for **scc**
RF	SMOTE	Very overfitted CM showing preference for **scc**
Extra-tree	Up-sample and down-sample to medium class	Overfitted CM showing preference for **bcc** and **mel**

Results - Binary

- Support Vector Classifier is selected as final binary classification model.
- SVC was less overfitted and it had higher performance on the validation set.

Average inference time per image: 0.35 seconds

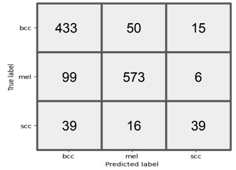
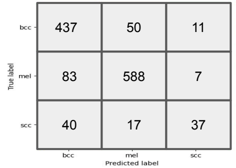
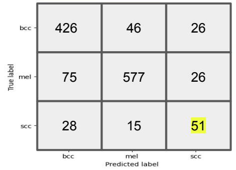
Classifier	F1 Score		Accuracy	
	Train	Val	Train	Val
SVC	**0.9359**	**0.8508**	**0.9359**	**0.8508**
Random Forest	1	0.8331	1	0.8332
XGBoost	0.9923	0.8373	0.9923	0.8374
Ensemble Model	0.993	0.8355	0.9929	0.8356
Gradient Boosting	0.8398	0.8102	0.8398	0.8102
K-nearest Neighbors	0.8807	0.8132	0.8806	0.8132
Adaboost	0.8171	0.8084	0.8171	0.8084
Logistic Regression	0.8097	0.8094	0.8097	0.8094

Results - Three class

- The best ensemble combination was selected based on the kappa score and a *balanced* confusion matrix.
- The weight vector was obtained empirically, to benefit the Random Forest classifier that showed better performance for **scc** classification

SVC | XGB | RF1 | RF2 | ERT

Average inference time per image: 0.36 seconds

Classifier	Kappa	CM
	Val	Val
Hard Voting	0.6788	
Soft Voting	0.7014	
Weighted Voting (soft) [1, 1, 4, 1, 1]	0.6976	

Conclusions

- Quite variate images coming from different datasets made it difficult to implement robust algorithms
- We had to resize the images (by half) when extracting features due to the large dataset and computational power constraints.
- We experimented we different complexity classifiers and obtaining their best training parameters (grid-search).
- For the binary problem we obtained an accuracy score of 0.8508 (validation set), thus concluding that the extracted features well characterize the two lesion classes.
- For the three class problem we chose to tune the ensemble method in favor of improving the classification of the three classes, getting a good kappa score and balanced confusion matrix.

References

[1] https://github.com/lcambero/skin_lesion_segmentation

[2] https://github.com/BlueDokk/Dullrazor-algorithm

[3] Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 31(6), 362–373. https://doi.org/10.1016/j.compmedimag.2007.01.003

[4] Hameed, N., Shabut, A. M., Ghosh, M. K., & Hossain, M. (2020). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. Expert Systems with Applications, 141, 112961. doi:10.1016/j.eswa.2019.112961

[5] D. N. H. Thanh, U. Erkan, V. B. S. Prasath, V. Kumar and N. N. Hien, "A Skin Lesion Segmentation Method for Dermoscopic Images Based on Adaptive Thresholding with Normalization of Color Models," 2019 6th International Conference on Electrical and Electronics Engineering (ICEEE), 2019, pp. 116-120, doi: 10.1109/ICEEE2019.2019.00030.

[6] Hameed, N., Ruskin, A., Abu Hassan, K., & Hossain, M. (2016). A comprehensive survey on image-based computer aided diagnosis systems for skin cancer. 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA). doi:10.1109/skima.2016.7916221

[7] Samsudin, S. S., Arof, H., Harun, S. W., Abdul Wahab, A. W., & Idris, M. Y. (2022). Skin lesion classification using multi-resolution empirical mode decomposition and local binary pattern. PLOS ONE, 17(9). doi:10.1371/journal.pone.0274896

[8] A. Sáez, J. Sánchez-Monedero, P. A. Gutiérrez and C. Hervás-Martínez, "Machine Learning Methods for Binary and Multiclass Classification of Melanoma Thickness From Dermoscopic Images," in IEEE Transactions on Medical Imaging, vol. 35, no. 4, pp. 1036-1045, April 2016, doi: 10.1109/TMI.2015.2506270.

[9] Gopinath, J., & John, M. (2015). Feature Extraction by Gabor Filter and Classification of Skin Lesion using Support Vector Machine.

This project was developed in the beautiful city of Girona, Cataluña

Not in Spain