

Computer Aided Diagnosis



- DERMOSCOPIC
- HISTOPATHOLOGY

Authors:

Patricia Cabanillas Silva
Zohaib Salahuddin

INTRODUCTION

- **Skin Challenge**

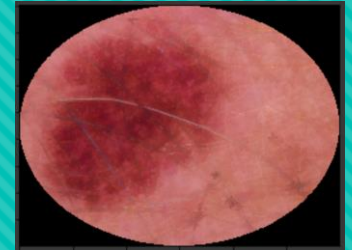
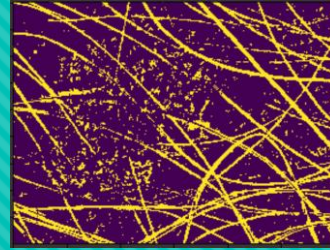
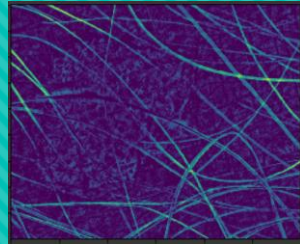
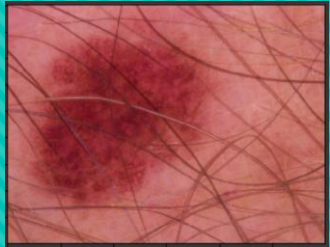
- Pre-Processing
- Feature Extraction
- Classifiers
- Final Result

- **Histopathology Challenge**

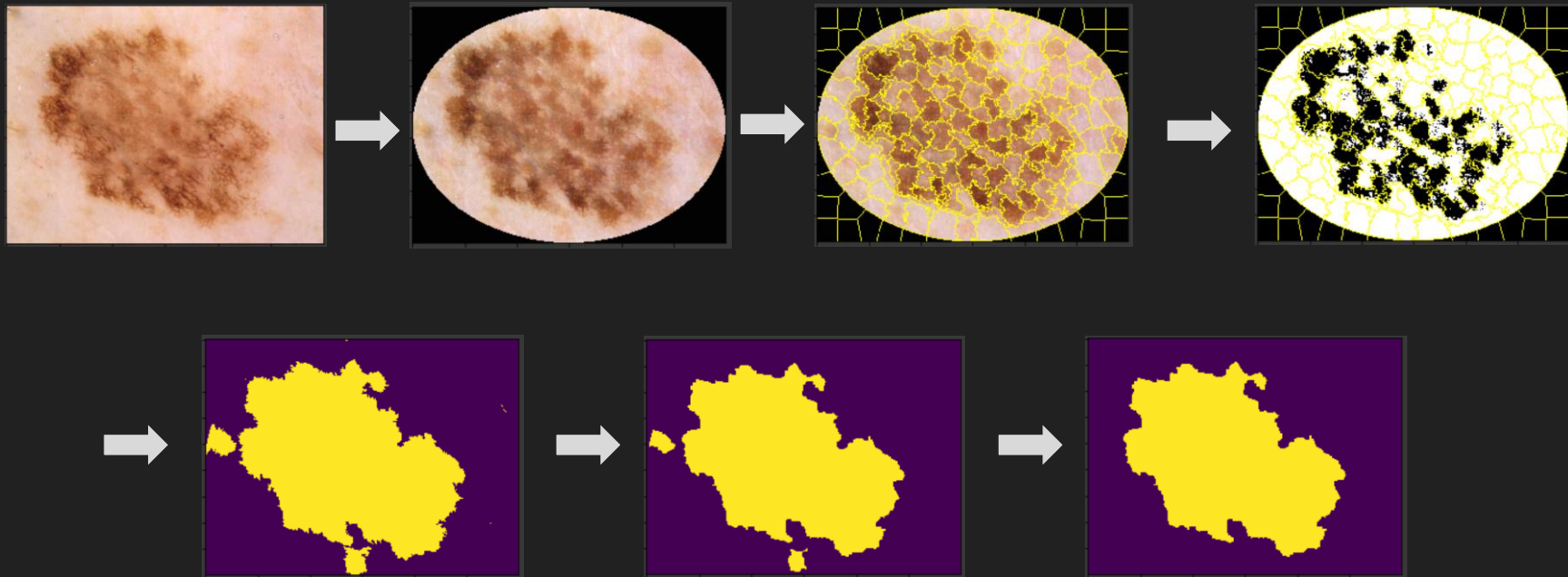
- Pre-Processing
- Feature Extraction
- Classifier
- Final Result

SKIN CHALLENGE

PRE-PROCESSING



Super Pixel Segmentation



SKIN CHALLENGE

- **Local Descriptors:**

- SURF

- **Shape Descriptors:**

- Hu moments
- Hog Features
- Circularity
- Perimeter
- Moments
- Zenrike Moments

- **Color Descriptors:**

- Histogram (HSV, RGB)
- Histogram (Mean and Standard Deviation)
- Skew
- Kurtosis

- **Texture Descriptors:**

- Haralick
- GLCM
- LBP

Dermoscopic : SURF and HOG Features

Surf with Sparse Coding

No of Words	200	400	600	800	1000
Accuracy	61.11	63.43	69.82	66.21	66.54

Surf with Dense Coding:
Spacing 10px

No of Words	200	400	600	800	1000
Accuracy	71.34	75.77	76.11	74.18	74.33
			PCA (200) = 75.81		

Surf with Dense Coding:
Spacing 20px

No of Words	200	400	600	800	1000
Accuracy	71.63	74.64	76.02	75.23	75.54
			PCA (200) = 75.29		

HOG Features:

Pixels Per Cell/Cells Per Block	(8,8)/(2,2)	(16,16)/(2,2)
Accuracy = (PCA = 200 Components)	74.28	73.78

SKIN CHALLENGE

	SVM C = 10
skew_f_w	0.693
skew_f	0.689
kurt_f_w	0.692
kurt_f	0.659
std_f_w	0.745
std_f	0.739
fd_histogram_g	0.803
fd_histogram_l	0.758
fd_histogram_l8	0.785
fd_histogram_g8	0.815

	SVM C = 10
fd_haralick_g	0.799
fd_haralick_lo	0.788
lbp_g	0.728
lbp_l	0.748
fd_hu_moments_l	0.670
desc	0.588
glcm_f	0.739
glcm_f_w	0.649

Annotations:

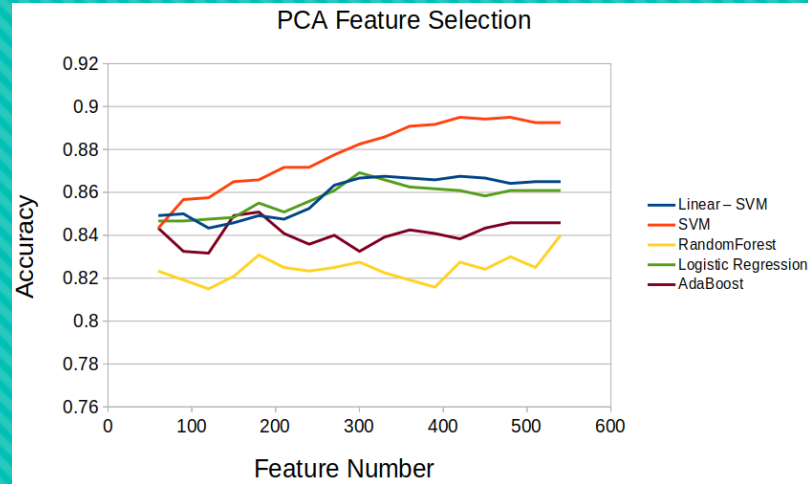
- f = features
- w = whole image
- g = grayscale
- l = color image

CHOSEN FEATURES AFTER TRIALS:

Skew, Kurtosis, Haralick, Std and Mean, Histogram
Hu moments, LBP Histogram

SKIN CHALLENGE

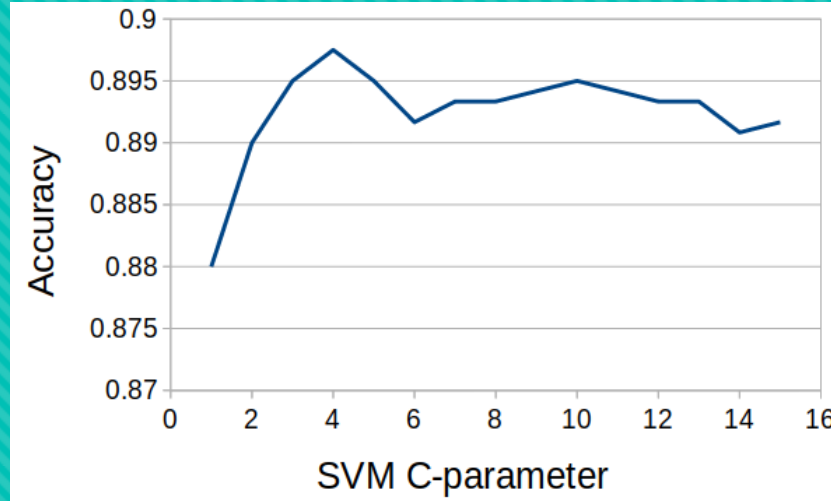
- **Classifiers:**
 - Linear SVM
 - SVM
 - Random Forest
 - Logistic Regression
 - AdaBoost



The maximum accuracy reached was **92.3** for training and **89.5** for validation, when **SVM** is used with 480 features.

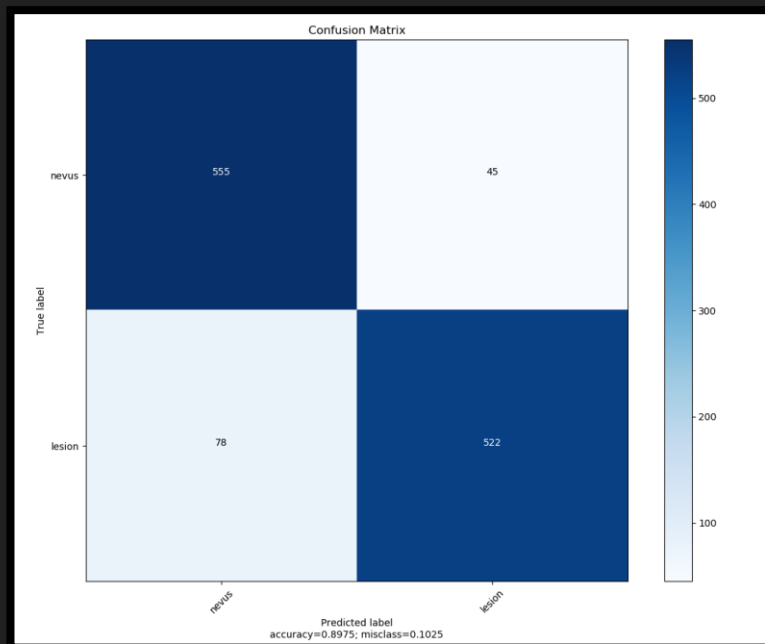
SKIN CHALLENGE

- **FINAL RESULTS:**
 - Fitting in c-parameter of SVM classifier
 - C - Parameter = 4, gives an accuracy of **89.75**



For the testing set, **SVM** Classifier with c-parameter = 4, was trained with the training and validation sets.

Dermoscopic: Confusion Matrix

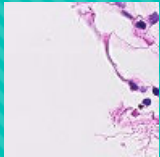
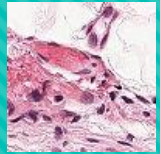
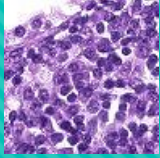


	precision	recall	f1-score	support
class 0	0.88	0.93	0.90	600
class 1	0.92	0.87	0.89	600
micro avg	0.90	0.90	0.90	1200
macro avg	0.90	0.90	0.90	1200
weighted avg	0.90	0.90	0.90	1200

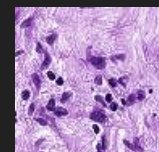
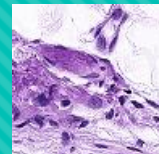
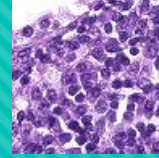
HISTOPATHOLOGY CHALLENGE

PRE-PROCESSING

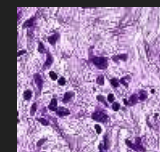
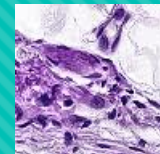
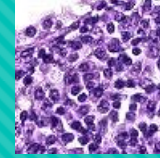
TARGET



Vahane - Normalization



CLAHE



HISTOPATHOLOGY : Feature Extraction

- **Shape Descriptors:**

- Hu moments
- Hog Features

- **Texture Descriptors:**

- Haralick
- GLCM
- LBP

- **Color Descriptors:**

- Histogram (HSV, RGB, LAB)
- Histogram (Mean and Standard Deviation)
- Skew
- Kurtosis

- **Local Descriptors:**

- SIFT
- SURF

Classifiers:

- **SVM Linear Kernel**
- **SVM Rbf Kernel**

- **Random Forests**
- **Adaboost**

HISTOPATHOLOGY : SIFT and HOG Features

Surf with Sparse Coding

No of Words	200	500	1000	1500	2000
Accuracy	67.23	70.58	72.08	73.55	74.88

Surf with Dense Coding:
Spacing 10px

No of Words	200	500	1000	1500	2000
Accuracy	70.44	74.33	76.82	77.17	78.91

Surf with Dense Coding:
Spacing 20px

No of Words	200	500	1000	1500	2000
Accuracy	71.23	73.44	75.22	77.22	77.11

HOG Features:

Pixels Per Cell/Cells Per Block	(8,8)/(2,2)	(16,16)/(2,2)
Accuracy = (PCA = 200 Components)	76.32	74.77

HISTOPATHOLOGY : Feature Analysis

	GLCM	HARALICK	Histogram	LBP Histogram	Gabor Features	HuMoments Features	SURF (200 PCA Components)	HOG
UnNormalized Linear SVM	0.756	0.771	0.792	0.773	0.808	0.68	0.767	0.71
UnNormalized Rbf SVM	0.772	0.783	0.813	0.793	0.819	0.692	0.781	0.763
UnNormalized Random Forest	0.712	0.743	0.773	0.743	0.778	0.69	0.743	0.73
UnNormalized AdaBoost Classifier	0.73	0.76	0.765	0.763	0.787	0.738	0.751	0.7544
Normalized Linear SVM	0.746	0.774	0.781	0.785	0.801	0.663	0.741	0.754
Normalized Rbf SVM	0.769	0.771	0.809	0.776	0.805	0.697	0.773	0.761

CHOSEN FEATURES AFTER TRIALS:

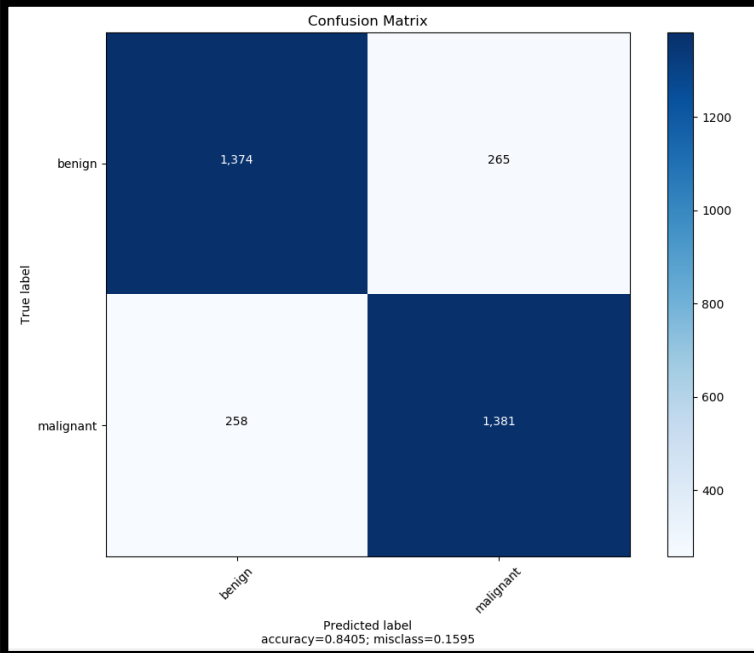
Haralick, GLCM, Histogram, LBP Histogram, Gabor Features

HISTOPATHOLOGY : Final Results

	Training Accuracy	Validation Accuracy	Mean 5 Fold Cross Validation Accuracy
UnNormalized Linear SVM	87.13	84.08	86.7
UnNormalized Rbf SVM	89.9	83.2	88.5
UnNormalized Random Forest	85.6	80.8	84.5
UnNormalized AdaBoost Classifier	87.74	81.5	84.9
Normalized Linear SVM	85.59	82.25	85.66
Normalized Rbf SVM	88.78	82.8	87.783

For the testing set, **Linear** Classifier was chosen for classifying on the basis of trained features.

Histopathology: Confusion Matrix



	precision	recall	f1-score	support
class 0	0.84	0.84	0.84	1639
class 1	0.84	0.84	0.84	1639
accuracy			0.84	3278

Conclusions

- For Dermoscopic Images, We were able to achieve **89.75 Percent Accuracy** on the Validation Set using the rbf SVM Classifier.
- For Histopathology Images, We were able to achieve **84.1 Percent Accuracy** on the Validation Set using the linear SVM Classifier. The mean cross validation accuracy was however, **88 Percent**.

Learning Outcomes:

- Evaluating Performance of different classifiers.
- Development of Feature Engineering Skills.
- Developing whole Machine Learning Pipeline of the project.