



**DEPARTMENT OF COMPUTER &
SOFTWARE ENGINEERING
COLLEGE OF E&ME, NUST, RAWALPINDI**



EC - 312 - Digital Image Processing
Semester Project

SUBMITTED TO:

**Dr. Usman Akram
LE Sundas**

SUBMITTED BY:

**Haadin Zaman
Reg # 00000351350
DE-42 (C&SE)-B**

**M. Hadi
Reg # 00000347360
DE-42 (C&SE)-B**

**M. Mamoon Khan
Reg # 00000337715
DE-42 (C&SE)-B**

**M. Faisal Iqbal
Reg # 00000335416
DE-42 (C&SE)-B**

Submission Date: 4-6-2023

Segmentation & Classification of Histopathological Skin Images Using Traditional Image Processing Techniques










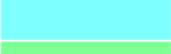
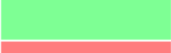

Objective

The objective of the project is to analysis the performance of a non-melanoma skin cancer classification model using conventional image processing techniques.

Dataset

The dataset provided was annotated and classified into three groups ‘BCC’, ‘IEC’ and ‘SCC’. The dataset also contained segmented mask of the image for a ground truth for each image. The cells of the slide were divided into 12 classes and assigned an RGB color value represented in the mask.

The dataset contained 1500 images and 1500 corresponding masks, the images and their corresponding masks were dived into training and testing sets with 1200 total images in the training set and 300 in the testing set with their masks divede equally for all three classes.

Tissue Type	Code	RGB Values	Color
Glands (sebaceous and sweat)	GLD	108, 0, 115	
Inflammation	INF	145, 1, 122	
Hair Follicle	FOL	216, 47, 148	
Hypodermis	HYP	254, 246, 242	
Reticular Dermis	RET	181, 9, 130	
Papillary Dermis	PAP	236, 85, 157	
Epidermis	EPI	73, 0, 106	
Keratin	KER	248, 123, 168	
Background	BKG	0, 0, 0	
Basal Cell Carcinoma	BCC	127, 255, 255	
Squamous Cell Carcinoma	SCC	127, 255, 142	
Intra-epidermal Carcinoma	IEC	255, 127, 127	

Methodology

Training

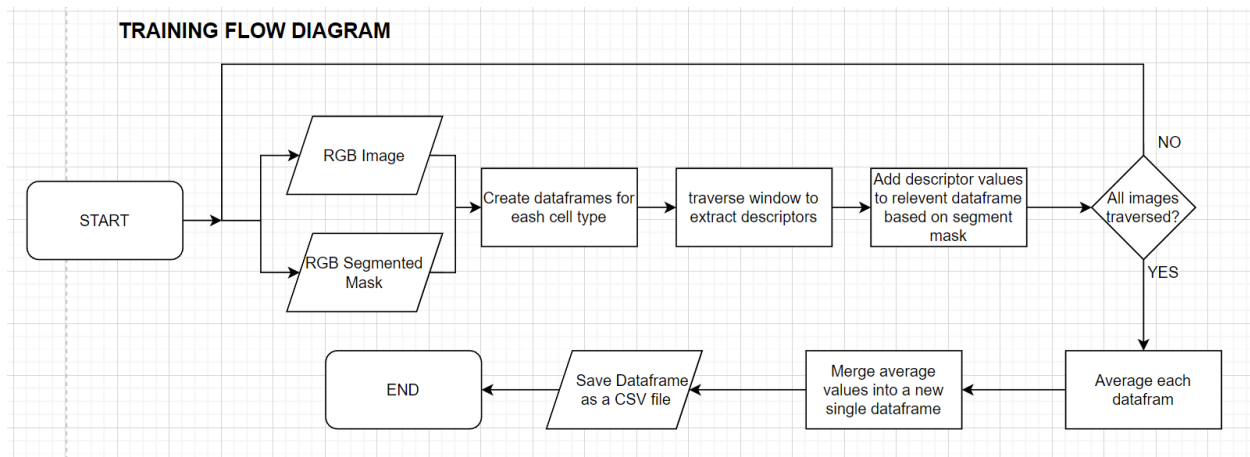
Training refers to the analysis of the given annotated data to decide on what features are to be selected which allow for the best representation of the different classes. Empirically it was found that texture descriptors prove to be a great measure to distinguish each class.

The selection of texture descriptors was extracted from a gray level correlational matrix (GLCM) applied to a localized path of the image knowing the assigned cell type from the give mask.

All descriptors were applied to a localized window to corelate obtained descriptor values with the cell types. The selected feature set was as follows:

- Contrast
- Dissimilarity
- Homogeneity
- Energy
- Correlation
- Entropy
- Mean
- Variance
- Sobel Edge Magnitude

The descriptor values were all collected for each cell type into individual pandas dataframes and finally the averages of each feature for each cell type were collected and merged into a single pandas dataframe to be saved as a CSV file and used to test and classify the testing dataset.



Mean descriptor values

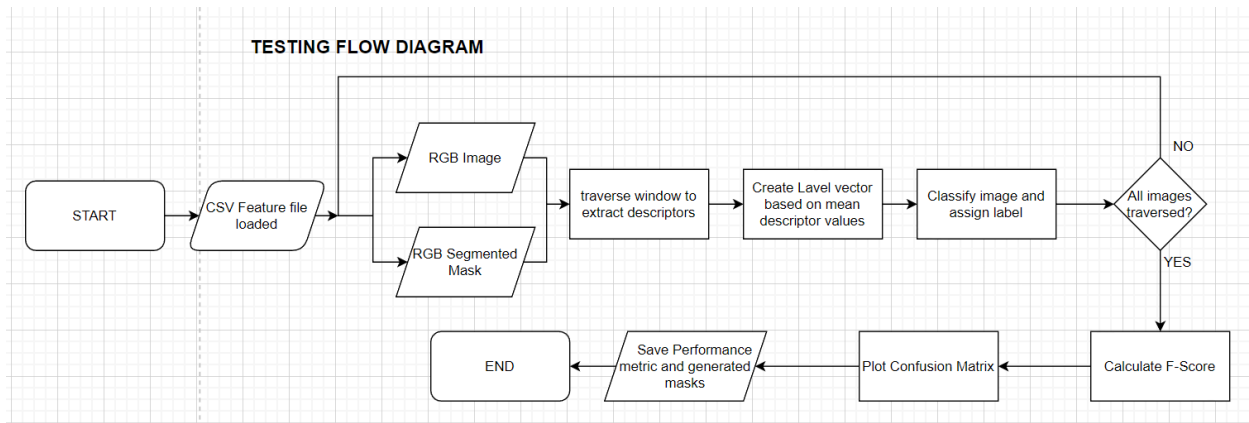
	contrast	dissimilarity	homogeneti	energy	correlation	entropy	mean	variance	sobel_mag
GLD	1475.467317	26.680000	0.081655	0.068057	0.267622	4.844691	229.239620	1606.814804	120.077754
INF	2149.286958	32.172435	0.072399	0.066860	0.146774	5.440077	221.263209	2110.059537	132.977818
FOL	1603.040458	28.065045	0.070599	0.060298	0.396854	6.081814	209.150480	2867.555566	130.428848
HYP	465.528720	9.123829	0.472150	0.407937	0.274381	2.543736	247.423173	417.383870	49.160809
RET	794.515527	17.784903	0.154131	0.115396	0.267562	4.185718	236.955655	897.318192	87.525004
PAP	1022.225921	20.681485	0.082387	0.059944	0.184730	5.138645	230.918689	1165.293937	88.906314
EPI	1302.949400	25.273683	0.088125	0.071859	0.486592	6.177550	204.945445	3174.880559	120.632584
KER	943.642454	19.629043	0.203989	0.171953	0.435987	4.663790	223.134624	1982.990528	100.308342
BKG	66.961541	1.246297	0.889430	0.863584	0.696701	0.559659	253.811739	82.574948	6.929996
BCC	2067.936766	34.228622	0.053894	0.060481	0.247392	6.284800	201.809668	3231.314201	146.815269
SCC	1021.273409	21.359841	0.083352	0.062950	0.236839	5.434539	220.065128	1984.023755	93.657489
IEC	1562.624379	28.786510	0.057332	0.056276	0.231116	6.408473	205.521078	2621.481029	119.162289

Testing

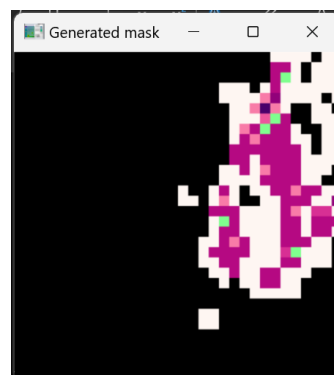
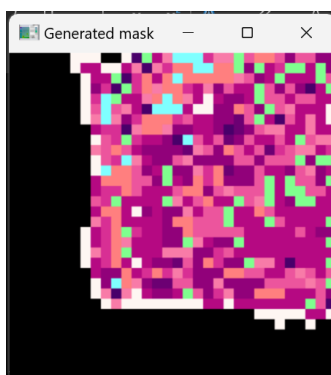
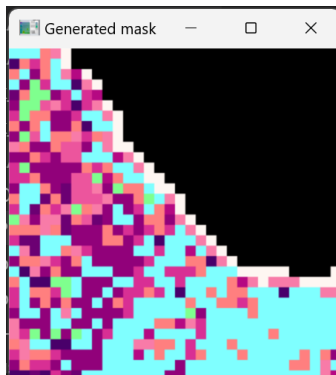
The Testing phase consists of reading the CSV file saved in the Training phase as a pandas dataframe containing the cell types on the row axis and their respective feature average values in the column axis.

The test image dataset was prepared at a multi-dimensional numpy array of each image stacked after each other in order and corresponding true labels in a separate numpy array in the same order.

- Each image is traversed with a window size of 8x8 pixels and a stride of 8 pixels.
- All previously mentioned feature descriptors are evaluated and compared with the mean feature values in the dataframe to assign a label for each feature it most probably belongs to, based on the nearest mean value
- This gives us a label vector for an image with labels based on each feature.
- The final classification is done based on the most occurring label in the label vector.
- Performance metrics are calculated.
- Confusion matrix is plotted.



Generated Masks

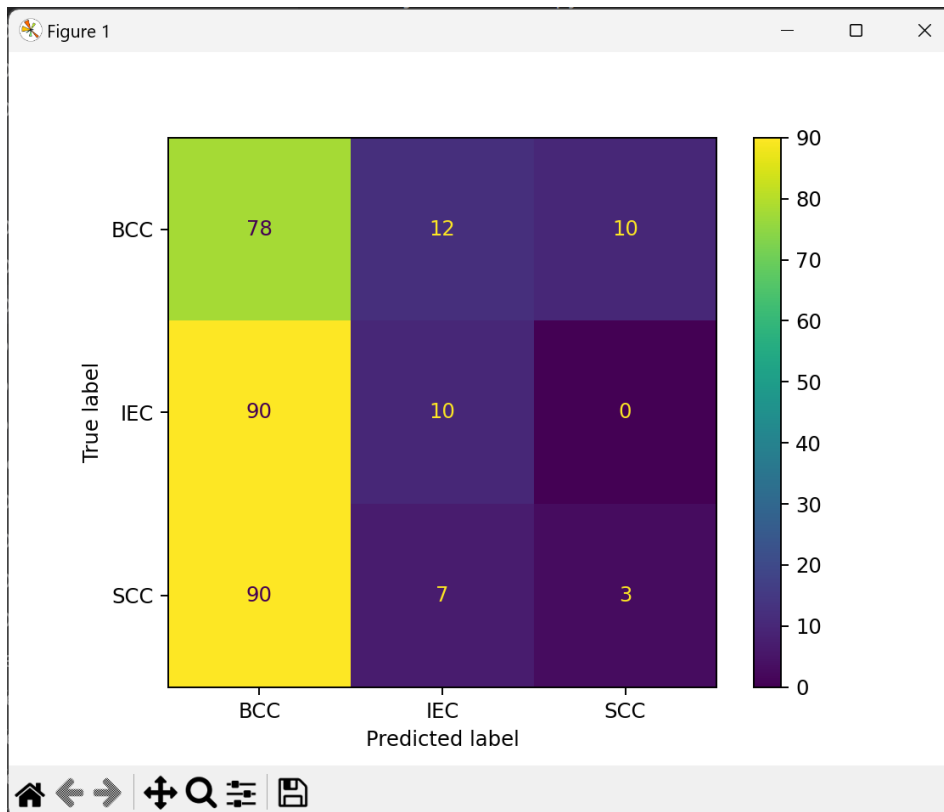


Performance metrics

F - score

```
-- F-Scores --  
  
BCC = 37.0  
IEC = 46.0  
SCC = 49.5  
  
-----
```

Confusion matrix



Conclusion

Based on the findings of this project Conventional image processing technique do provide capabilities to classify Non-Melanoma cancer types with ~50 % effectiveness but lack the

deep feature extraction possible using Machine Learning models which can learn and adapt after each result.

GitHub Repository

[Hayzee99/Non-Melenoma-Cancer-Detection-Using-Tradtional-Image-Processing \(github.com\)](https://github.com/Hayzee99/Non-Melenoma-Cancer-Detection-Using-Tradtional-Image-Processing)