Digital Image Processing
Complex Engineering Problem (End Term project)
(CLO4 -> PLO5)
Analysis of Histopathological Images

Non-melanoma skin cancers comprise over 90% of all skin cancer diagnoses, compared to melanoma which comprises approximately 4%. The most common forms of non-melanoma skin cancer include basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) which are among the most common forms of cancer in Australia and the United States. BCC comprises approximately 60% of all skin cancer diagnoses, with SCC comprising a further 30%. Another non-melanoma skin cancer is intraepidermal carcinoma (IEC), otherwise known as squamous carcinoma in situ or Bowen's disease.

Dataset:

The given dataset is divided into three major classes:

- BCC – Basal Cell Carcinoma
- IEC – Intra-epidermal Carcinoma
- SCC – Squamous Cell Carcinoma

500 patches of 256x256 dimensions each of the above classes have been shared with their relevant image and ground truth. The dataset in total contains 12 classes, namely: Glands (GLD), Inflammation (INF), Hair Follicles (FOL), Hypodermis (HYP), Reticular Dermis (RET), Papillary Dermis (PAP), Epidermis (EPI), Keratin (KER), Background (BKG), BCC, SCC, and IEC. Following RGB color codes are assigned to each class:

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	

You can download data using
https://drive.google.com/file/d/1qGGai4oB2jt1skkMIBEBLTmSjwukAaZQ/view?usp=share_link

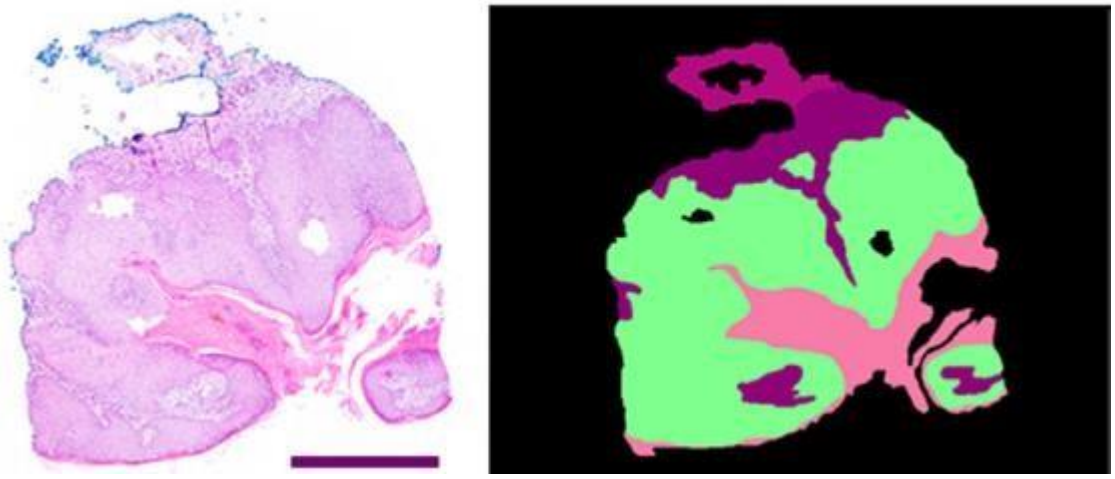
In this applied nature project, you have to **design and develop** a framework that is capable of following:

- Segmenting out all 12 kind of pixels
- Classifying an patch into one of the 03 given classes

This would require you to do a complete study about the problem and read relevant literature. After getting requisite knowledge about the problem, you need to use all the knowledge which you have gathered during this course along with your other mathematics and programming courses to build the complete solution.

Segmentation:

Segmentation of an object or pixels in the image is the process of identifying all the pixels from input image that belong to a particular object/class(12 in this problem). An example of segmentation is given below:



Classification:

Usually, the classification output can be represented in the one of the following ways:

- Either each of the output classes can be assigned a probability; the input image has 80% probability of belonging to BCC, 13% probability to belonging to IEC and only 7% probability of belonging to SCC.
- We can simply say that the class with the maximum probability is the class of the image (so, in this case, that is BCC).

Data Division:

From given data, you need to separate out 100 images from each class as validation or testing data for which you will report all performance parameters. You can use remaining images to design and develop your solution.

Performance Metrics:

Performance metrics tell us about the performance of the classifier that is used. The performance metrics to be calculated for the classification tasks are as follows:

- Overall accuracy
- Confusion matrix

Similarly, the performance metric to be used for segmentation is the Dice Coefficient (or F1 score) given by the formula:

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}.$$

For further discussion on the performance metrics, the following link provides a good explanation:

https://en.wikipedia.org/wiki/F-score **Submission:**

The results are to be computed for the validation subset (100 Images from each folder). Your submission on LMS should have following.

- Make a zip file including ○ a report containing description about your solution, flow diagram, tabular values of evaluation parameters and some sample outputs
  - ○ All code files
- All code and associated files are to be submitted via a public GitHub account as well.

Related Links:

The following provide a useful starting points for the different subparts of the project:
https://www.sciencedirect.com/science/article/abs/pii/S1361841520302796?via%3Dihub
https://github.com/topics/semantic-segmentation