Departamento Ingeniería
Eléctrica

Breast Cancer Recurrence Prediction using a Quantitative Evaluation of the Nuclear Pleomorphism

February 25, 2019

M.Sc. Ricardo Moncayo, Ph.D. Anne Martel, Ph.D. Eduardo Romero

cim@lab Computer Imaging & Medical Applications Laboratory

Pathological examination issues

- Due to the heterogeneity between the healthy tissue and the disease, most of the morphological indicators of prognosis are not quantified

Pathological examination issues

- Due to the heterogeneity between the healthy tissue and the disease, most of the morphological indicators of prognosis are not quantified
- The observations are susceptible to inter and intra observer variations

The Suitability of a Quantitative Estimation in Pathology Workflow

- Digital version of the whole glass slide and image analysis are using to improve the diagnosis

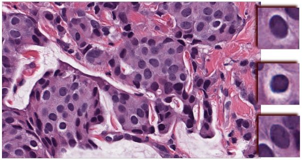
The Suitability of a Quantitative Estimation in Pathology Workflow

- Digital version of the whole glass slide and image analysis are using to improve the diagnosis
- The Quantification of morphological structures to determine the predisposition of a disease and to deliver healthcare

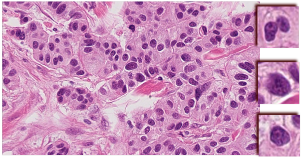
Nuclear Pleomorphism (NP) in Ductal carcinoma in situ (DCIS) - case of Study

Nuclear Pleomorphism

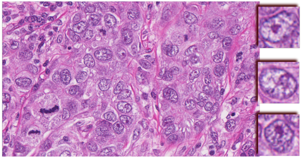
- The current research is on automatic strategies to quantify nuclear pleomorphism in breast cancer(NPBca).



grade 1



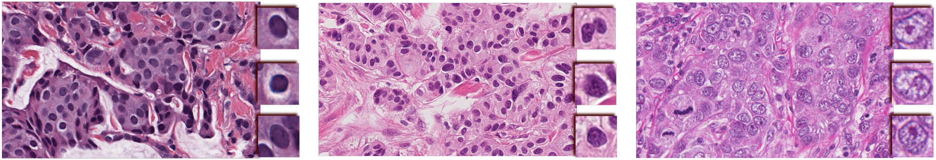
grade 2



grade 3

Nuclear Pleomorphism

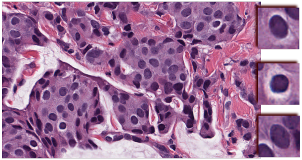
- The current research is on automatic strategies to quantify nuclear pleomorphism in breast cancer(NPBca).
- NPBca is an indicator of the aggressiveness of the disease



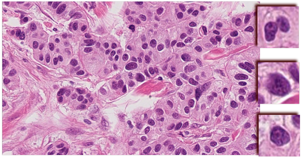
grade 1 grade 2 grade 3

Nuclear Pleomorphism

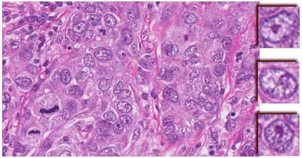
- The current research is on automatic strategies to quantify nuclear pleomorphism in breast cancer(NPBca).
- NPBca is an indicator of the aggressiveness of the disease
- NP has a low inter-observer agreement ($0.3 < \kappa < 0.5$)



grade 1



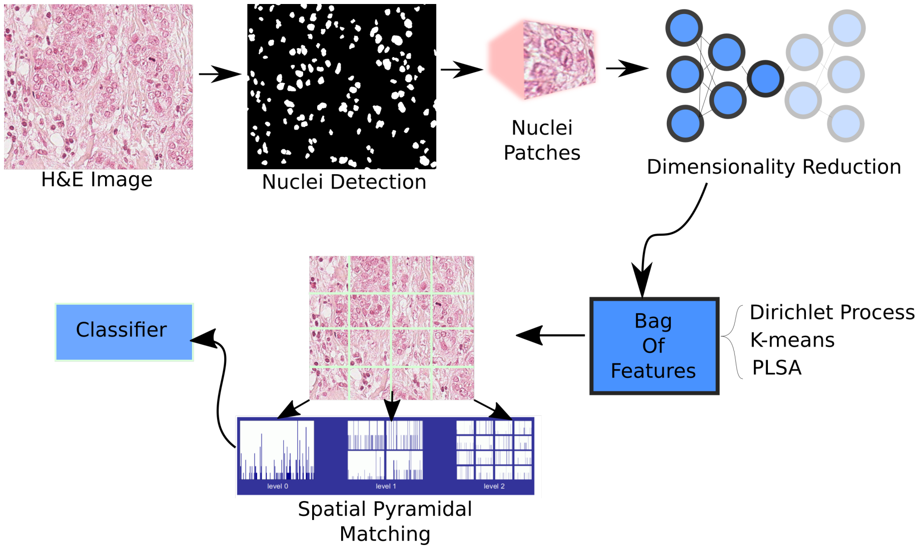
grade 2



grade 3

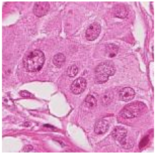
Proposed Framework

Spatial Pyramidal Characterization (SPYC)



Nuclei Detection

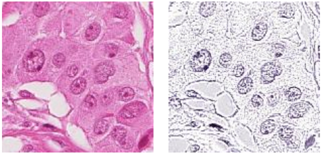
- Using a color deconvolution technique the Hematoxylin (H) and eosin (E) stains are estimated [1]



[1] A method for normalizing histology slides for quantitative analysis, Macenko et al.ISBI 2009

Nuclei Detection

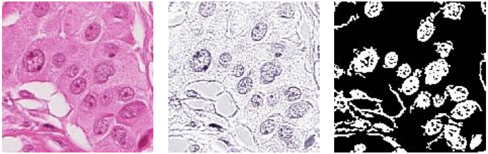
- Using a color deconvolution technique the Hematoxylin (H) and eosin (E) stains are estimated [1]



[1] A method for normalizing histology slides for quantitative analysis, Macenko et al.ISBI 2009

Nuclei Detection

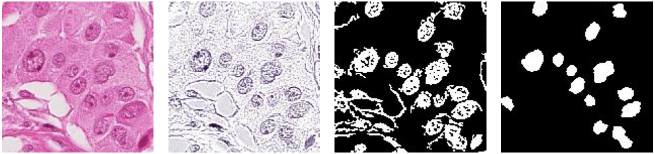
- Using a color deconvolution technique the Hematoxylin (H) and eosin (E) stains are estimated [1]
- Watershed algorithm is used to detect nuclei candidates in H stain



[1] A method for normalizing histology slides for quantitative analysis, Macenko et al.ISBI 2009

Nuclei Detection

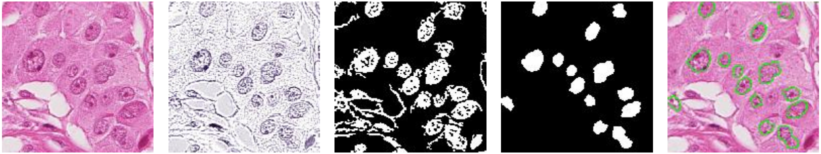
- Using a color deconvolution technique the Hematoxylin (H) and eosin (E) stains are estimated [1]
- Watershed algorithm is used to detect nuclei candidates in H stain
- Final candidates are found after morphological operations



[1] A method for normalizing histology slides for quantitative analysis, Macenko et al.ISBI 2009

Nuclei Detection

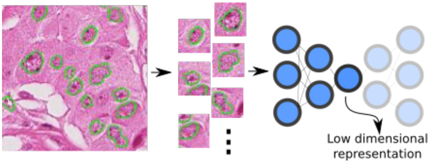
- Using a color deconvolution technique the Hematoxylin (H) and eosin (E) stains are estimated [1]
- Watershed algorithm is used to detect nuclei candidates in H stain
- Final candidates are found after morphological operations



[1] A method for normalizing histology slides for quantitative analysis, Macenko et al.ISBI 2009

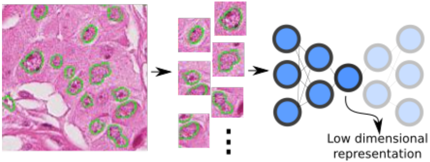
Nuclei Patches And Dimensionality Reduction

- A patch of $n \times n$ pixels around the candidate nucleus is extracted



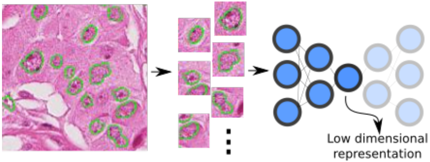
Nuclei Patches And Dimensionality Reduction

- A patch of $n \times n$ pixels around the candidate nucleus is extracted
- An autoencoder with three convolutional layers for encoding and 3 for decoding



Nuclei Patches And Dimensionality Reduction

- A patch of $n \times n$ pixels around the candidate nucleus is extracted
- An autoencoder with three convolutional layers for encoding and 3 for decoding
- The coded representation for each nucleus is a vector of 2048 dimensions (22.5 compression factor)



Bag of Features Building

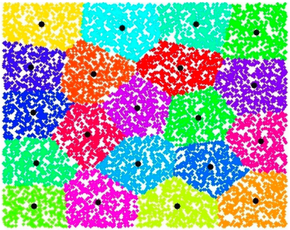


Figure: *K-means Algorithm*

- The multi-scale feature space is partitioned with a *k*-means algorithm

Bag of Features Building

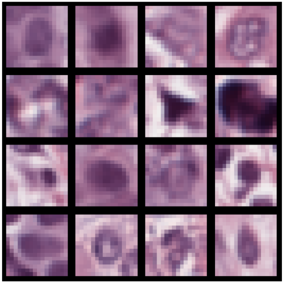


Figure: *Decoded Dictionary Representation*

- The multi-scale feature space is partitioned with a *k*-means algorithm
- Each centroid correspond to a visual word in the BoF

Bag of Features Building

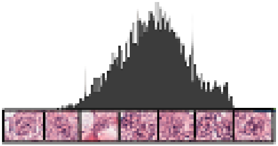


Figure: *Ocurrence representation*

- The multi-scale feature space is partitioned with a *k*-means algorithm
- Each centroid correspond to a visual word in the BoF
- Each new candidate from a train or test image is represented by some atom in the dictionary, building an histogram.

Spatial Pyramidal Matching

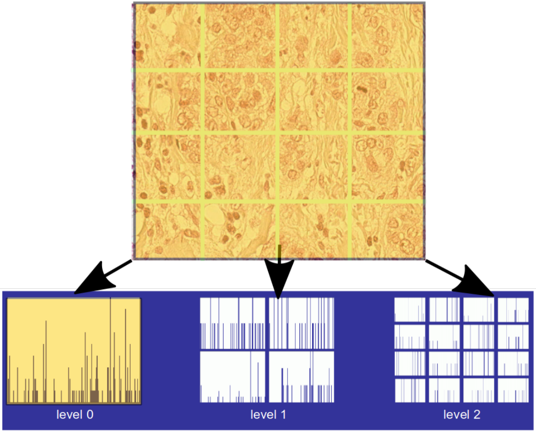
- Locally representation at several levels of resolution

Spatial Pyramidal Matching

- Locally representation at several levels of resolution
- The construction is as follow:

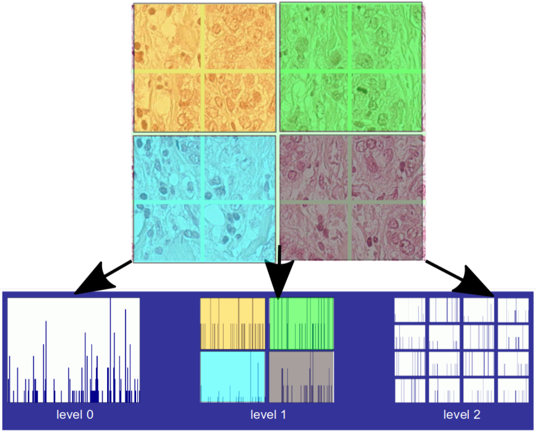
Spatial Pyramidal Matching

- Locally representation at several levels of resolution
- The construction is as follow:



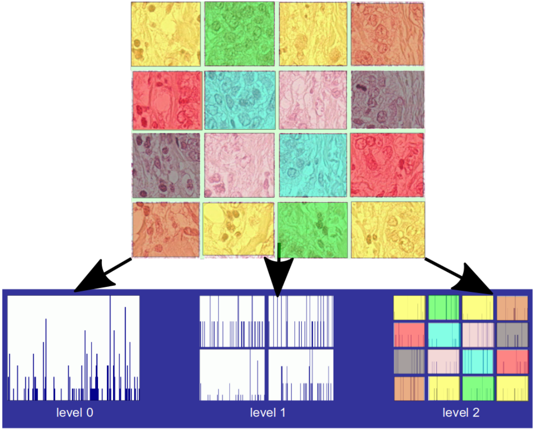
Spatial Pyramidal Matching

- Locally representation at several levels of resolution
- The construction is as follow:



Spatial Pyramidal Matching

- Locally representation at several levels of resolution
- The construction is as follow:

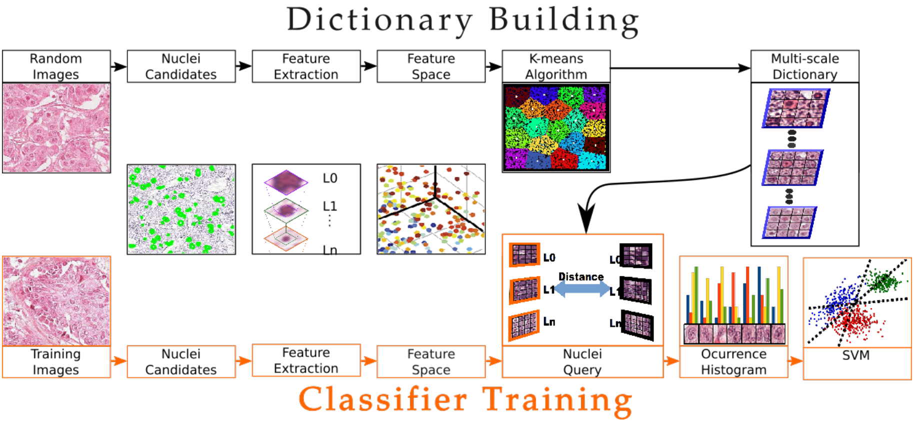


Spatial Pyramidal Matching

- Locally representation at several levels of resolution
- The construction is as follow:
- The histograms are concatenated in a single vector and feed the classifier

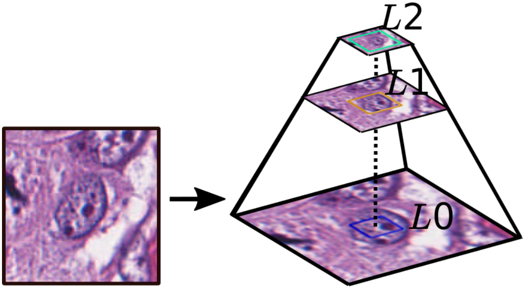
$$f(Y) = [Hist(l0), hist(l1), hist(l2)]$$

Bag of Features with Multiscale Descriptor Baseline Framework(BoF)



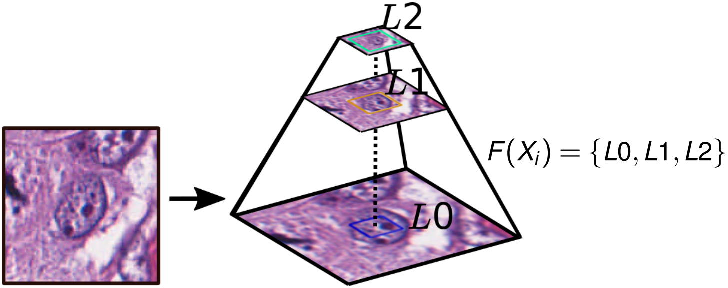
Multi-scale Feature Extractor

- The characterization of each nucleus candidate was performed by analyzing multiple scales

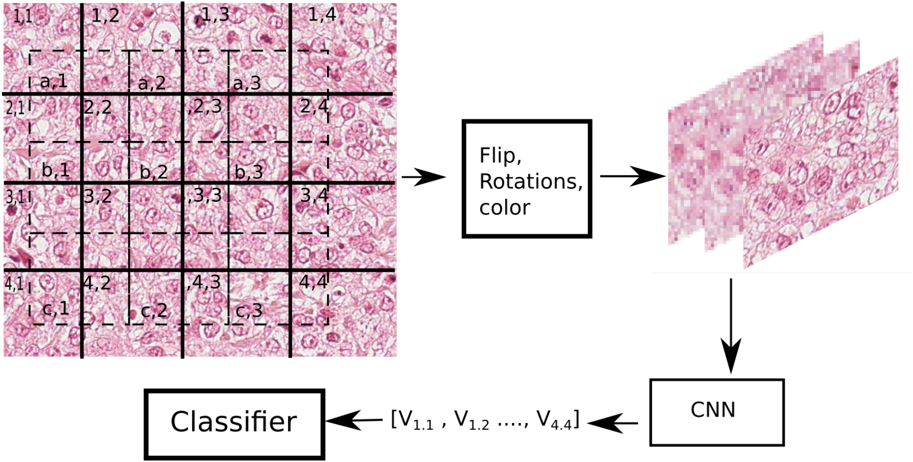


Multi-scale Feature Extractor

- The characterization of each nucleus candidate was performed by analyzing multiple scales
- The feature vector corresponds to the information from RGB patches which are concatenated along one dimension.



Convolutional Neural Networks Baseline Framework (CNN)



CNN Baseline Framework

- The Image is divided in several patches

CNN Baseline Framework

- The Image is divided in several patches
- A Data augmentation step is performed

CNN Baseline Framework

- The Image is divided in several patches
- A Data augmentation step is performed
- Different CNN are trained (VGG16, Alexnet, Resnet50, CNN) one at time

CNN Baseline Framework

- The Image is divided in several patches
- A Data augmentation step is performed
- Different CNN are trained (VGG16, Alexnet, Resnet50, CNN) one at time
- The activation output is used as a feature vector of the patch and then is concatenated with the subsequents to obtain a feature vector of the whole image

CNN Baseline Framework

- The Image is divided in several patches
- A Data augmentation step is performed
- Different CNN are trained (VGG16, Alexnet, Resnet50, CNN) one at time
- The activation output is used as a feature vector of the patch and then is concatenated with the subsequents to obtain a feature vector of the whole image
- The feature vector is used to feed a classifier

Experiments and Results

Experiments

- Public database - Mitos Atypia 14
- Patches of 72x72 pixels are extracted using the nuclei centroid as seed
- SPYA framework was evaluated also using each level independently and altogether

Dataset- Mitos Atypia 14

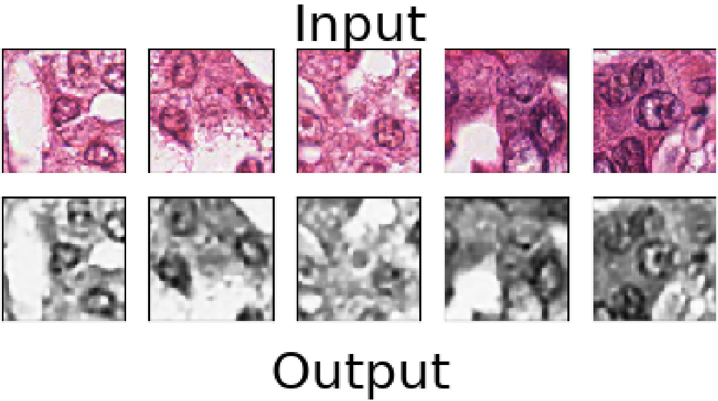
Original		After Augmentation		
Grade	#40X images	Total Patches 344x344 pixels	Train	Validation
NP 1	92	21,344	18,144	3,200
NP 2	900	22,500	19,096	3,404
NP 3	208	20,800	17,764	3,036

Table: *Training and validation - 11 cases for training*

Grade	# 40X images
NP 1	152
NP 2	252
NP 3	92

Table: *Test dataset - 5 cases*

Visual Performance of the Autoencoder



Autoncoder Reconstruction

Results using F1-Score

The F-score is used to evaluate the classification

$$F_\beta = (1+\beta^2).\frac{precision.recall}{(\beta^2.precision) + recall} \tag{1}$$

$\beta = 1$ is used this is the harmonic mean between precision and recall

Results - Proposed Framework

Task	SPYA	SPYA L0	SPYA L1	SPYA L2
3 Vs. 1 and 2	0.578	**0.630**	0.565	0.530
1 Vs. 2 and 3	**0.624**	0.564	0.520	0.489
1 Vs 2	0.473	**0.597**	0.542	0.573
2 Vs 3	0.560	0.603	0.606	**0.632**
1 Vs 3	0.623	**0.716**	0.639	0.632

Table: *Results using a coded representation of the nuclei*

Results of CNN Baseline

	CNN (1024dim)	CNN (3dim)	VGG16	Alexnet	Resnet50
3Vs.1 and 2	0.635	0.523	0.629	0.493	**0.685**
1Vs.2 and 3	**0.676**	0.645	0.564	0.581	0.573
1 Vs 2	0.637	**0.664**	0.453	0.623	0.592
2 Vs 3	0.570	0.567	0.544	0.619	**0.670**
1 Vs 3	0.666	**0.672**	0.620	0.649	0.563

Results of BoF multiscale descriptor

Experiment/S.Patch	32x32 L2	20x20 L2
1 vs 2 and 3	**0.593**	0.434
3 vs 1 and 2	**0.642**	0.47
1 vs 2	**0.7128**	0.47
1 vs 3	0.660	**0.7029**
2 vs 3	**0.6759**	0.5813

Table: *Results Bag Of Features*

Results of BoF multiscale descriptor

Experiment/S.Patch	32x32 L2	20x20 L2
1 vs 3	0.660	**0.7029**
2 vs 3	**0.6759**	0.5813

Table: *Results Bag Of Features*

The BoF with the multiscale descriptor presents a better performance that the SPYA!, a new experiment of this without autoencoder is evaluated...

Results without Autoencoder

	BOF PY	BOF PY L0	BOF PY L1
3 Vs. 1 and 2	0.624	0.593	0.583
1 Vs. 2 and 3	**0.692**	0.642	0.678
1 Vs 2	0.650	**0.713**	0.660
2 Vs 3	0.722	0.660	**0.729**
1 Vs 3	0.628	0.675	**0.740**

Results without Autoencoder

	BOF PY	BOF PY L0	BOF PY L1
3 Vs. 1 and 2	0.624	0.593	0.583
1 Vs. 2 and 3	**0.692**	0.642	0.678
1 Vs. 2	0.650	**0.712**	0.660

Then, we obtain a better model of NP using the spatial pyramidal representation aproach.

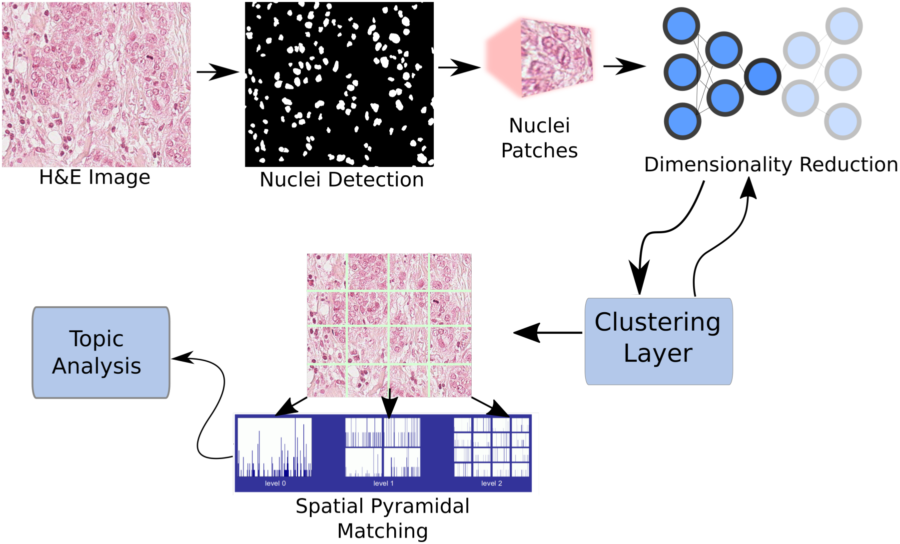
Final Remarks

- The Slightly variation between the NP grades and the high tissue variability makes it challenging to find an adequate model
- The CNN model could be improved if take advantage of the spatial information
- The auto-encoder representation and vocabulary construction maybe are not the most accurate

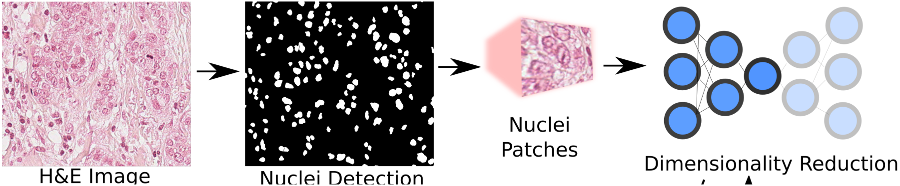
Current Work

- To determine a better cluster representation
- To evaluate the frameworks in a new database
- To determine if there exists a relationship between NP grading and the cancer recurrence (Future work)

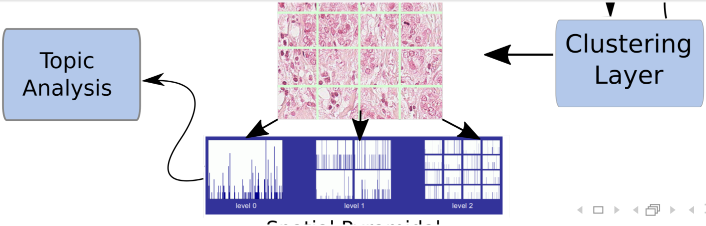
New proposal



New proposal

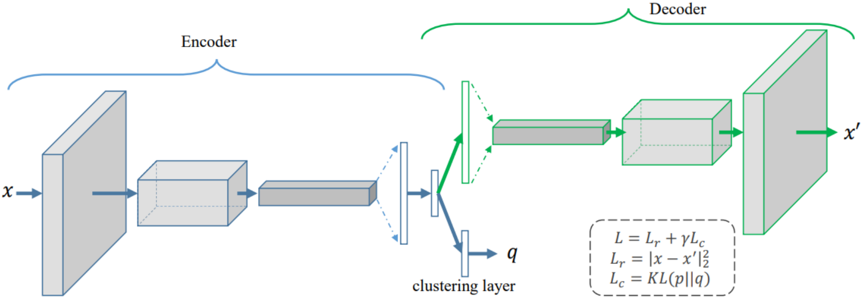


Two main diferences of the previous aproach are presented: 1) The clustering step, 2) A topic representation



Clustering Step

- The proposed deeplearning architecture[2] minimize the reconstruction loss of the autoencoder and the clustering loss
- Intends to preserve the structure of the data



[2] Guo et al., ICONIP17

New Dataset

- 300 cases of DCIS with recurrence time from Sunnybrook Health Institute
- 306 marked regions from 24 cases grading by an expert pathologist

Grading	# Marked Areas
Grade 1	135
Grade 2	132
Grade 3	39
Total Cases	**24**

Thank you...