

Departamento Ingeniería
Eléctrica

Presentation Title

Presentation Subtile

February 22, 2019

M.Sc. Ricardo Moncayo

Advisor: Ph.D. Eduardo Romero



Pathological examination issues

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

Pathological examination issues

- Due to the heterogeneity between the normal tissue and the disease most of the morphological indicators of prognosis are not quantified
- Subjective observations are susceptible to inter and intra observer variations **Imagenes con kappa**

The Suitability of a Quantitative Estimation in Pathology Workflow

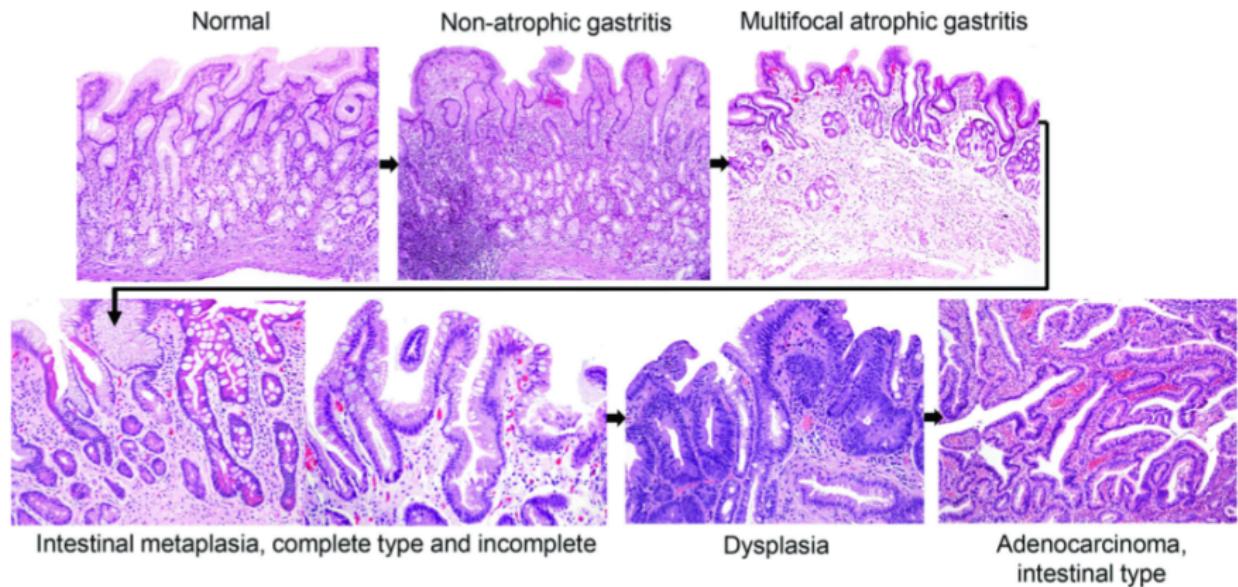
- Digital version of whole glass slide and image analysis are using to improve the diagnosis

The Suitability of a Quantitative Estimation in Pathology Workflow

- Digital version of 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 **imagenes??**

Quantification of Premalignant Lesions of the Gastric Mucosa - 1st case of Study

Changes of the Normal Mucosa to Adenocarcinoma

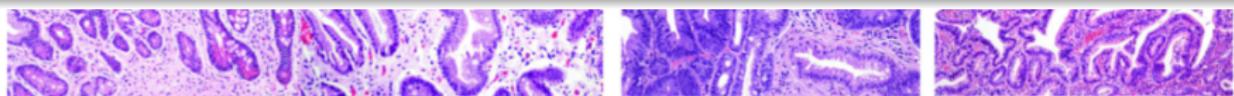


Correa's Cascade of Gastric Carcinogenesis;1976

Changes of the Normal Mucosa to Adenocarcinoma



Some studies suggest that the cancer progression could not follow this sequence!



Intestinal metaplasia, complete type and incomplete

Dysplasia

Adenocarcinoma,
intestinal type

Correa's Cascade of Gastric Carcinogenesis;1976

The Carcinogenesis Cascade Main Concerns

- Diffuse type of gastric cancer are not related with atrophy or metaplasia (Lauren classification)

¹ Stolte M, Zeitschrift für Gastroenterologie, 1992

² Filipe ML, International Journal of Cancer, 1994

³ Hattori T., Cancer 1986

⁴ Takizawa T, Stomach and intestine, 1998

The Carcinogenesis Cascade Main Concerns

- Diffuse type of gastric cancer are not related with atrophy or metaplasia (Lauren classification)
- In the non-atrophic gastritis foci of intestinal metaplasia are found¹

¹ Stolte M, Zeitschrift für Gastroenterologie, 1992

² Filipe ML, International Journal of Cancer, 1994

³ Hattori T., Cancer 1986

⁴ Takizawa T, Stomach and intestine, 1998

The Carcinogenesis Cascade Main Concerns

- Diffuse type of gastric cancer are not related with atrophy or metaplasia (Lauren classification)
- In the non-atrophic gastritis foci of intestinal metaplasia are found¹
- Metaplasia grade I and II have been show no risk for cancer progression²¹

¹ Stolte M, Zeitschrift für Gastroenterologie, 1992

² Filipe ML, International Journal of Cancer, 1994

³ Hattori T., Cancer 1986

⁴ Takizawa T, Stomach and intestine, 1998

The Carcinogenesis Cascade Main Concerns

- Diffuse type of gastric cancer are not related with atrophy or metaplasia (Lauren classification)
- In the non-atrophic gastritis foci of intestinal metaplasia are found¹
- Metaplasia grade I and II have been show no risk for cancer progression²
- Dysplasia are not found in micro-early carcinomas (*Diameter < 5mm*)³

⁴

¹ Stolte M, Zeitschrift für Gastroenterologie, 1992

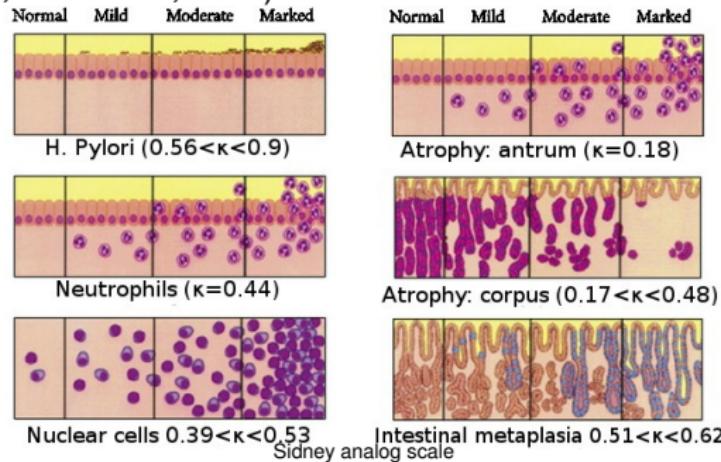
² Filipe ML, International Journal of Cancer, 1994

³ Hattori T., Cancer 1986

⁴ Takizawa T, Stomach and intestine, 1998

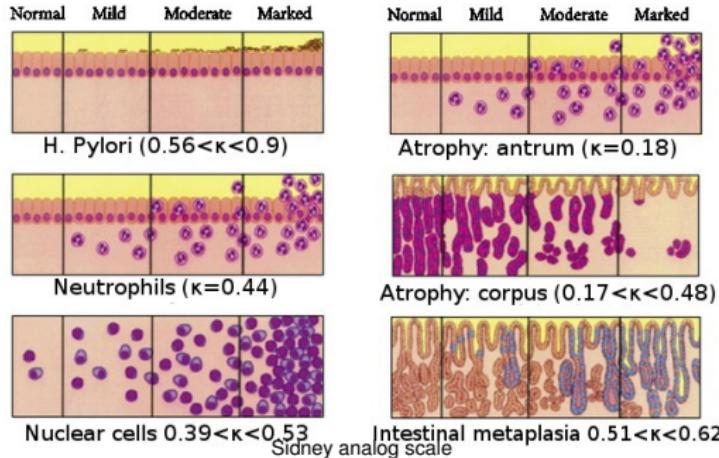
Inter-observer Variability

- A highly disagreement between experts reported using kappa coefficient (Aydin,2003;Offerhaus,1999)



Inter-observer Variability

- Kappa coefficient measure the agreement between observers, $\kappa = 0$ corresponds to a bad agreement; $\kappa = 1$ corresponds to a good agreement



Processes involved in gastric cancer incidence are not understood, and the current grading scales are expert dependent, A Careful and reliable measurement of histological features may prevent further malignant outcomes^{5,6,7,8}

⁵Tepes,2004

⁶Sipponen P,1994

⁷Meining A,2001

⁸Correa P., 1992

Hypothesis

An accurate quantification of the tissue structures will allow finding a suitable characterization of the disease

Objectives

General Objective:

To formulate an automatic model to find and quantify patterns in gastric premalignant lesions in histological images of the gastric mucosa to the different degrees of atrophy and intestinal metaplasia

Contribution

- The proposal consists in highlighting the importance of an exact measurement in gastric premalignant lesions and the possibility of using this in the pathology workflow.

Expected results

- Different strategies to establishes the stage in PGC lesions
- A computer-aided system to be applied by a pathologist in the workflow

Materials and Methods

Materials

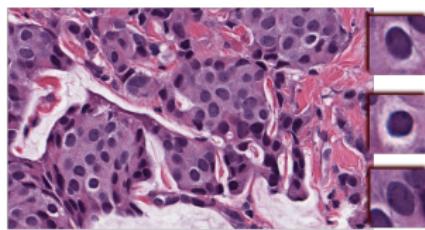
- TCGA Gastric cancer open database with 771 whole slide images and 443 of this with their histological description
- A temporal cohort of histological images with PGC from Urkunina 5000 project
- 30 GCa cases from Universidad Nacional de Colombia

Nuclear Pleomorphism (NP) in Ductal carcinoma in situ (DCIS) - 2nd case of Study

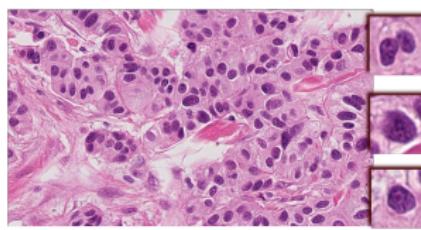
Nuclear Pleomorphism (NP) - The Suitability of a Quantitative Estimation

Nuclear Pleomorphism in BC

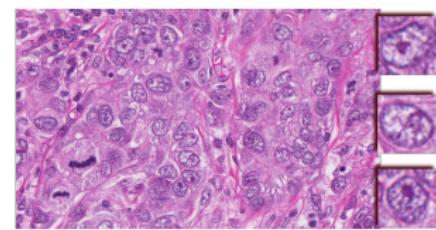
- Researching in automatic quantification strategies for nuclear pleomorphism in breast cancer(NPBca).



grade 1



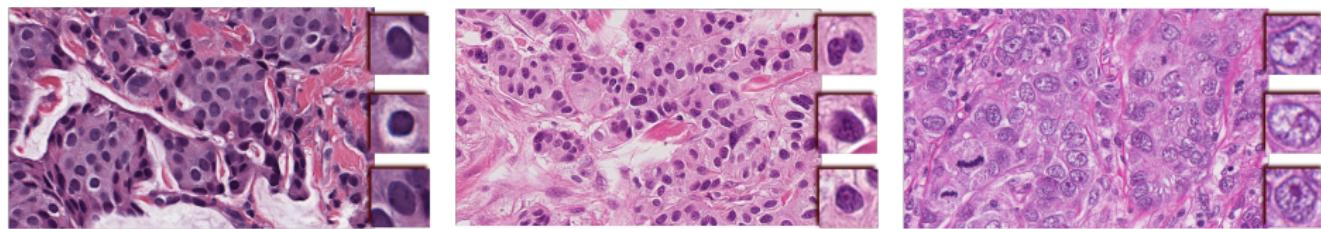
grade 2



grade 3

Nuclear Pleomorphism in BC

- Researching in automatic quantification strategies for nuclear pleomorphism in breast cancer(NPBca).
- NPBca is an indicator of the aggressiveness of the disease



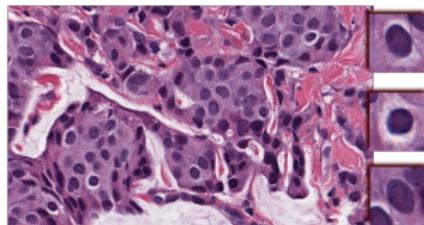
grade 1

grade 2

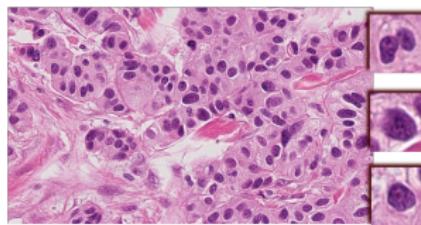
grade 3

Nuclear Pleomorphism in BC

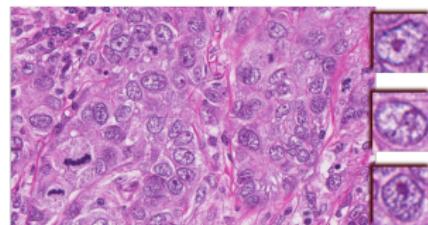
- Researching in automatic quantification strategies for nuclear pleomorphism in breast cancer(NPBca).
- NPBca is an indicator of the aggressiveness of the disease
- A low inter-observer agreement ($0.3 < \kappa < 0.5$)



grade 1

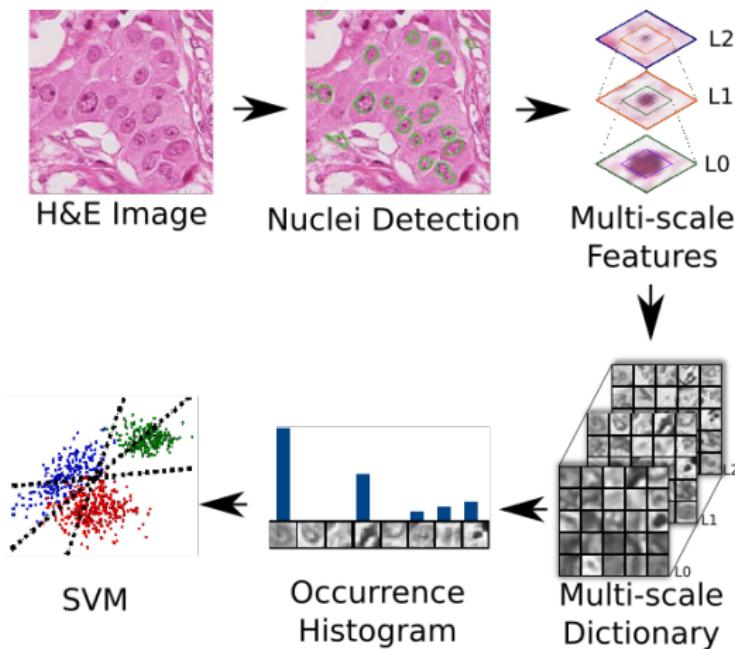


grade 2

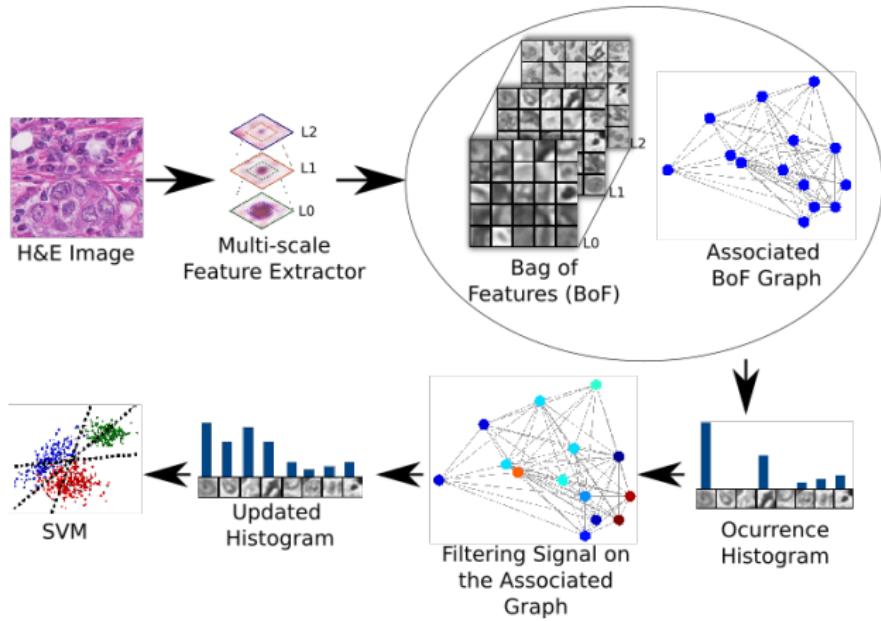


grade 3

Previous: A Bag of Features Based Strategy



Previous: A Bag of Features onto a Graph Strategy



Research Products

- Moncayo, R., Romo-Bucheli, D., Romero, E. **A Grading Strategy for Nuclear Pleomorphism in Histopathological Breast Cancer Images Using a Bag of Features (BoF)**. In Progress in Pattern Recognition, Image Analasys, Computer Vision and Applications, Pages 75-82, november 2015. ISBN 978-3-319-25751-8
- Romo-Bucheli, D., Moncayo, R., Cruz, A., Romero, E., **Identifying histological concepts on basal cell carcinoma images using nuclei based sampling and multi-scale descriptors**. In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), April 2015, pages 1008-1011, ISSN 1945-7928
- Moncayo, R., Romo-Bucheli, D., Arias, V., Romero, E., **Scoring Nuclear pleomorphism using a visual BoF modulated by a graph structure**, In the 13th International Symposium on Medical Information Processing and Analysis, October 2017

The Next Step

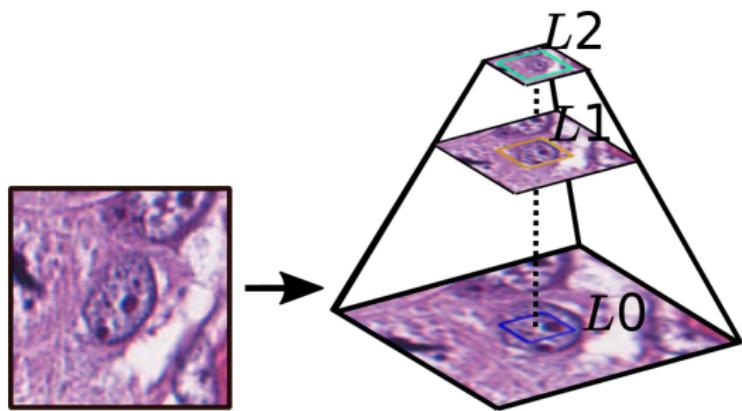
The use of nuclear pleomorphism grade would be correlated with the patient outcome resulting in a reliable measure for the patient

Materials

Coming soon...

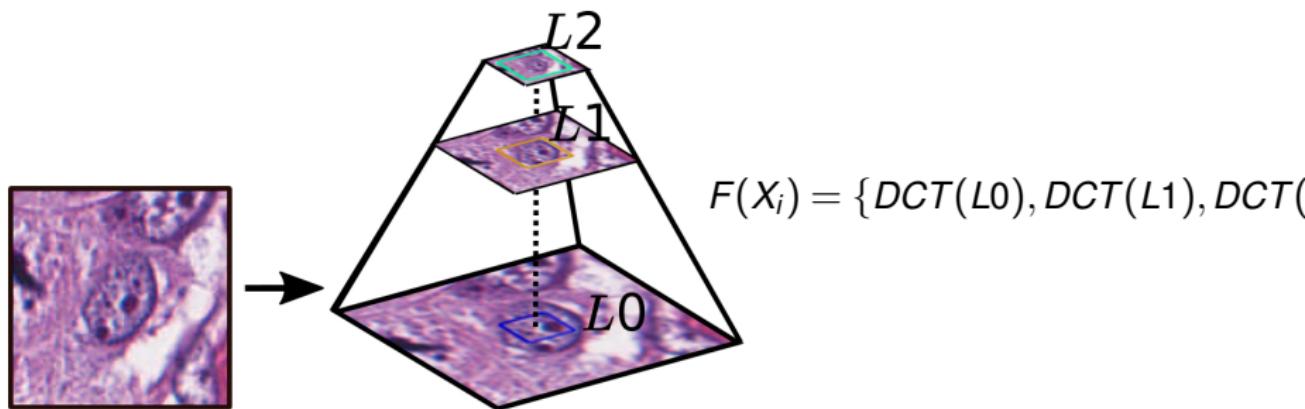
Multi-scale Features

- Each nucleus candidate is represented by multiple scales characterized by the discrete cosine transform (DCT) using enough coefficients to reconstructing the original image
- The coefficients from different scales are concatenated into a single feature



Multi-scale Features

- Each nucleus candidate is represented by multiple scales characterized by the discrete cosine transform (DCT) using enough coefficients to reconstructing the original image
- The coefficients from different scales are concatenated into a single feature



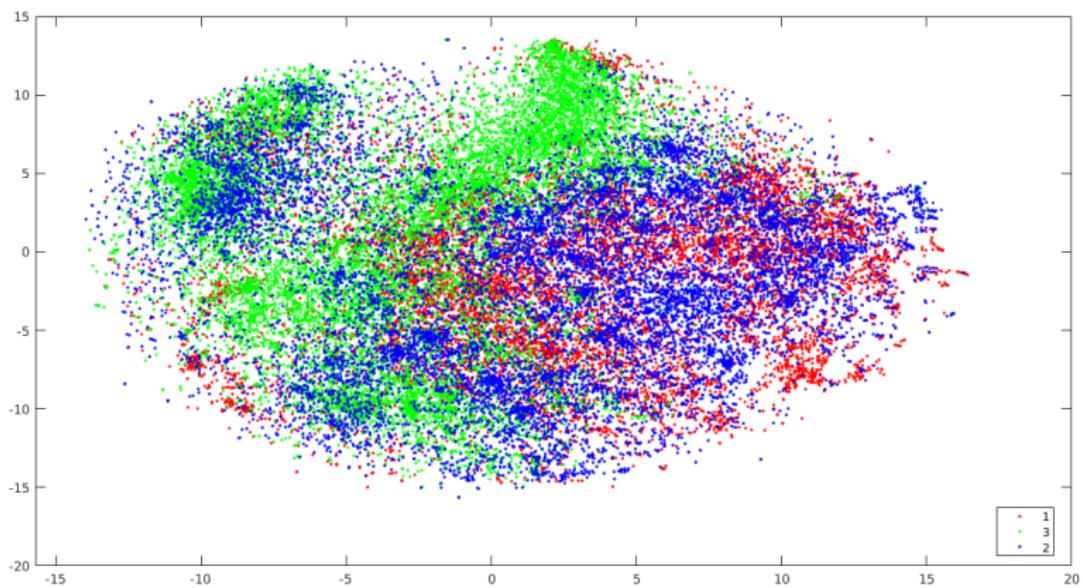
Probabilistic Latent Semantic Analysis(PLSA)

- Each multiscale representation is used as a **document** for build a PLSA model to find some topics

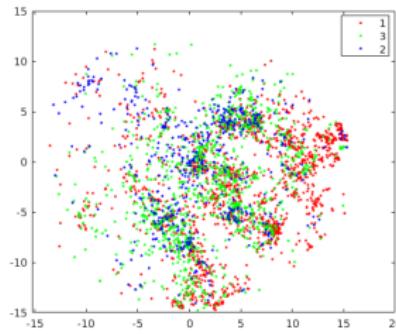
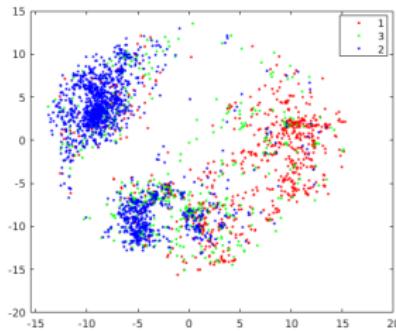
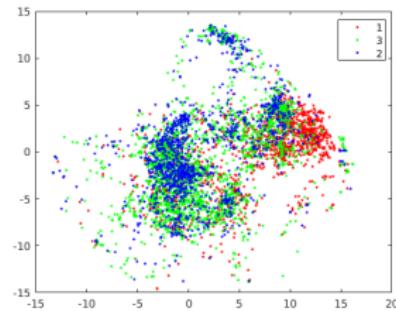
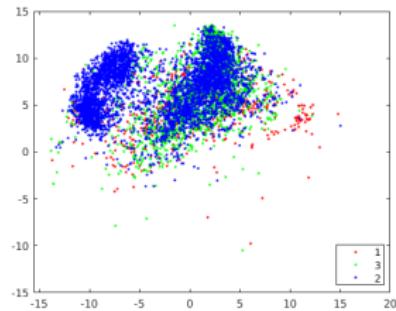
Probabilistic Latent Semantic Analysis(PLSA)

- Each multiscale representation is used as a **document** for build a PLSA model to find some topics
- T-sne is used to visualize the original space and the PLSA generated

Visual Results - Original Space



Visual Results - Space by topic



Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

Director: Ph.D Eduardo

Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

- Digital anatomy through images

Director: Ph.D Eduardo

Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

Director: Ph.D Eduardo Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

- Digital anatomy through images
- Digital Pathology

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

Director: Ph.D Eduardo Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

- Digital anatomy through images
- Digital Pathology
- Motion and bio-signal analysis

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

Director: Ph.D Eduardo Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

- Digital anatomy through images
- Digital Pathology
- Motion and bio-signal analysis
- Medical information systems and technology

Research Group



The Computer Imaging and Medical Applications Laboratory (CIM@LAB) is a research group devoted to solve pertinent medical problems by combining artificial vision and machine learning techniques.

Classification: A1

Director: Ph.D Eduardo Romero

Professors: 2

Doctoral Students: 9

Master Students: 19

- Digital anatomy through images
- Digital Pathology
- Motion and bio-signal analysis
- Medical information systems and technology

Finished PH.D. thesis: 9

Finished M.Sc. thesis:
46

International indexed journals: 37

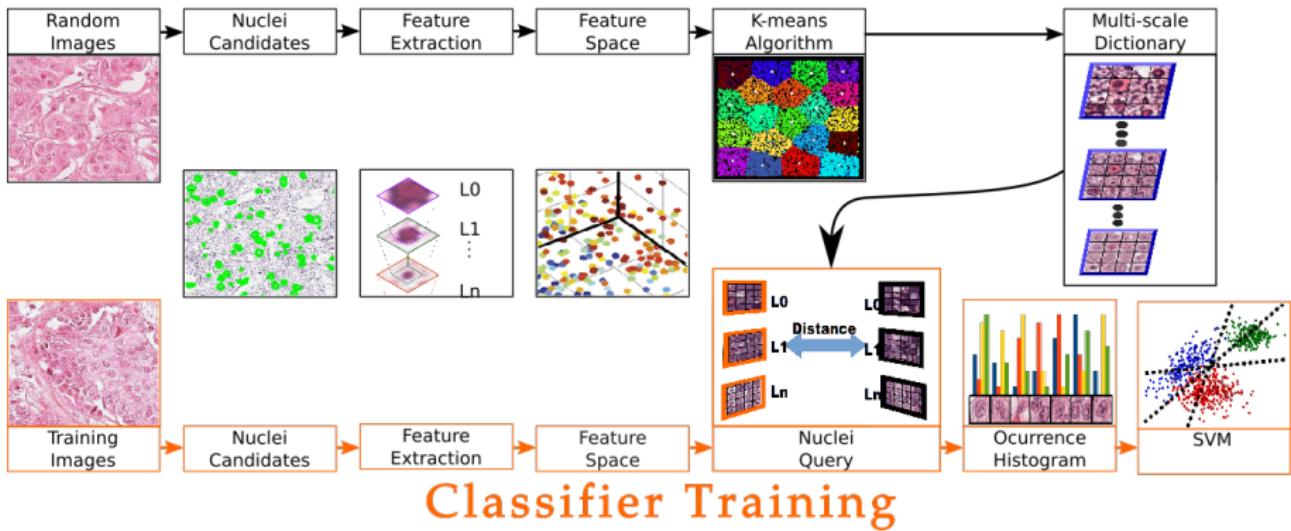
International peer reviewed conferences:
100

Books & chapters: 8

Colombian journals: 15
Colombian conferences: 21

First Aproach

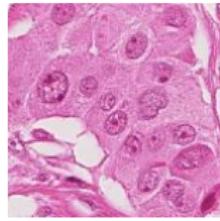
Dictionary Building



Classifier Training

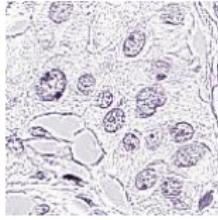
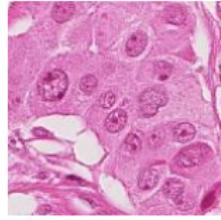
Nuclei Detection

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



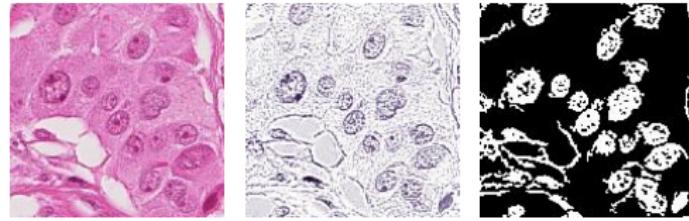
Nuclei Detection

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



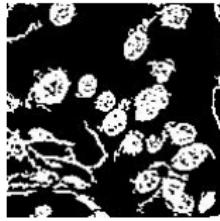
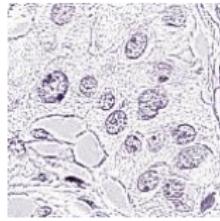
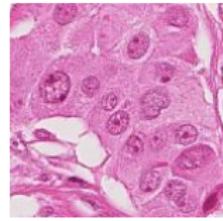
Nuclei Detection

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



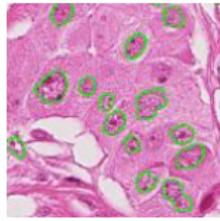
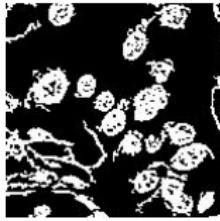
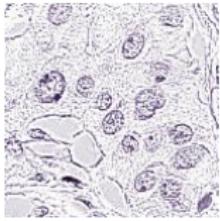
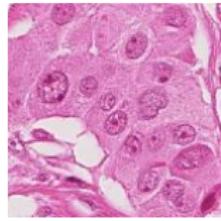
Nuclei Detection

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



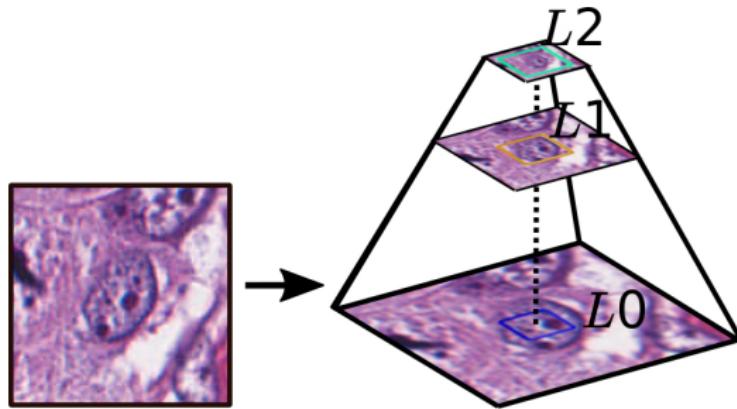
Nuclei Detection

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



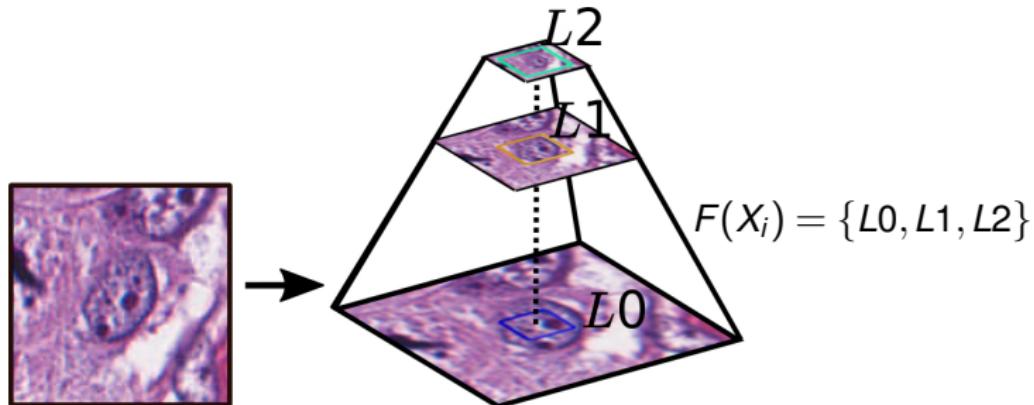
Multi-scale Feature Extractor

- 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.



Results F1-Score

The F-score is used to evaluate the classification

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

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

Dataset - Mitos Atypia 14

Grade	#40X images
NP 1	92
NP 2	900
NP 3	208

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*

Results

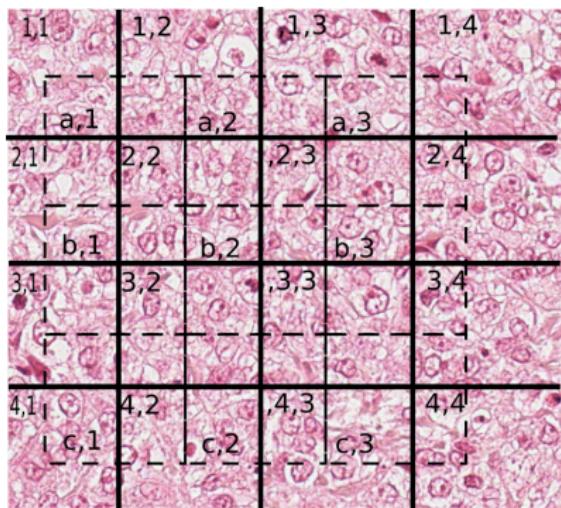
Experiment/S.Patch	32x32 L2	20x20 L2
1 vs 2 and 3	0.593	0.434
3 vs 1 and 2	0.642	0.47
1 vs 2	0.7128	0.47
1 vs 3	0.660	0.7029
2 vs 3	0.6759	0.5813

Table: *Results Bag Of Features*

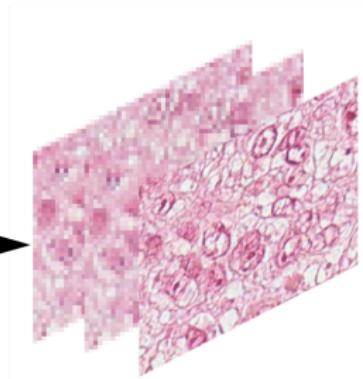
Current Work

- **To enhance the automatic nuclear pleomorphism grading**
- To find out if there is a relationship between the nuclear grade quantification with the cancer recurrence

CNN Framework



Flip,
Rotations,
color



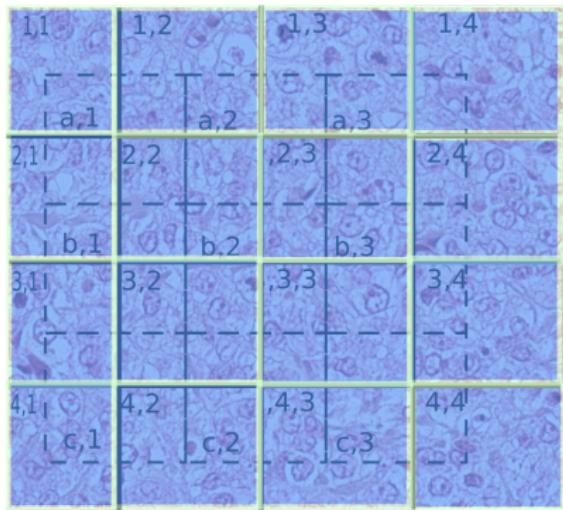
CNN

[$V_{1,1}, V_{1,2}, \dots, V_{4,4}$]

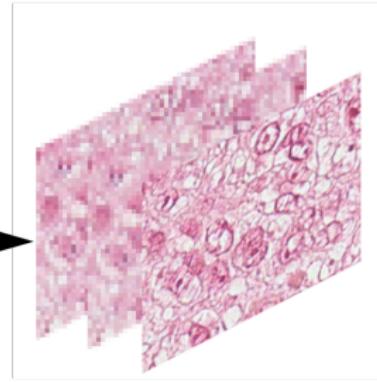


Classifier

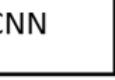
CNN Framework



Flip,
Rotations,
color



CNN



Classifier

[V_{1,1}, V_{1,2} ..., V_{4,4}]

Dataset- Mitos Atypia 14

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*

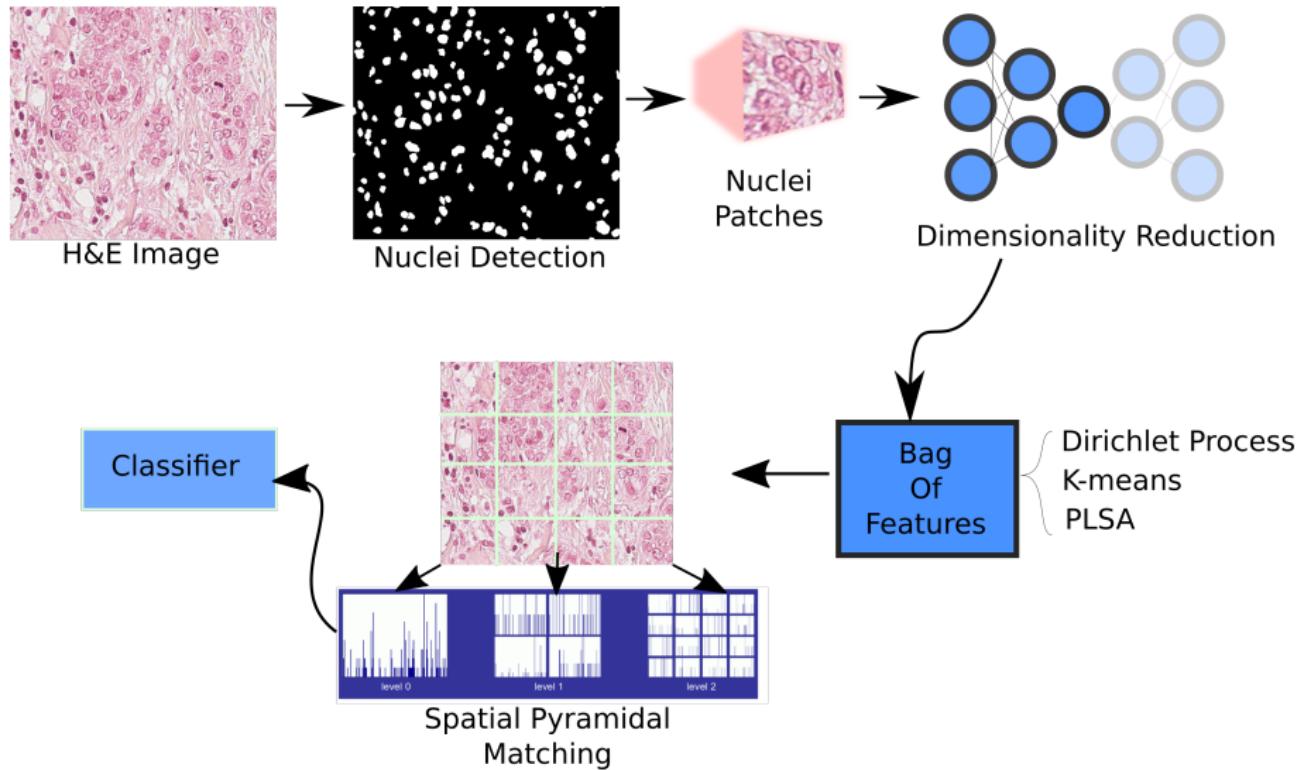
Grade	# 40X images
NP 1	152
NP 2	252
NP 3	92

Table: *Test dataset - 5 cases*

Results

	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

Spatial Pyramidal Characterization

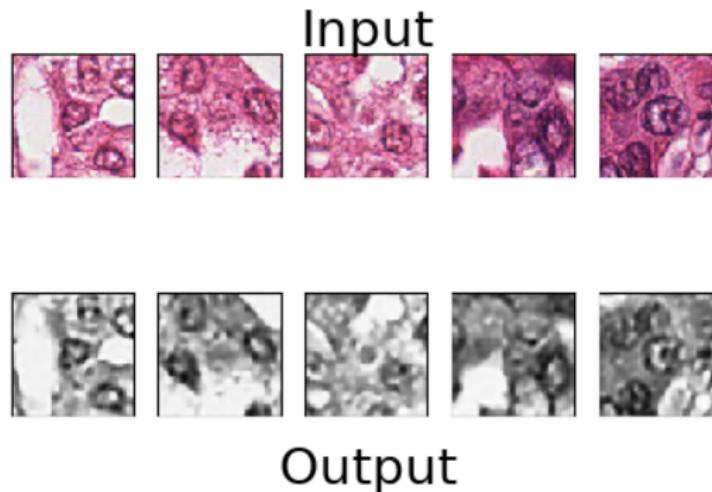


Nuclei Detection

- Color deconvolution to find hematoxylin channel (Macenko et al.)
- Watershed Algorithm to find connected components at different scales (Veta et al.)
- Patches are extracted using the nuclei centroid as seed (72x72 pixels)

Dimensionality Reduction- Convolutional autoencoder

- 3 convolutional layer for encoding and 3 for decoding
- 2048 dimensions in coded representation for nuclei (22.5 compression factor)
- 6300 dim feature vector (L0,L1,L2)



Autoncoder Reconstruction

Results

	PS.	PS L0	PS L01	PS L2	BOF PY	BOF PY L0	BOF PY L1
3 Vs. 1 and 2	0.578	0.630	0.565	0.53	0.624	0.593	0.583
1 Vs. 2 and 3	0.624	0.564	0.520	0.489	0.692	0.642	0.678
1 Vs 2	0.473	0.597	0.542	0.573	0.650	0.713	0.660
2 Vs 3	0.560	0.603	0.606	0.632	0.722	0.660	0.729
1 Vs 3	0.623	0.716	0.639	0.632	0.628	0.675	0.740

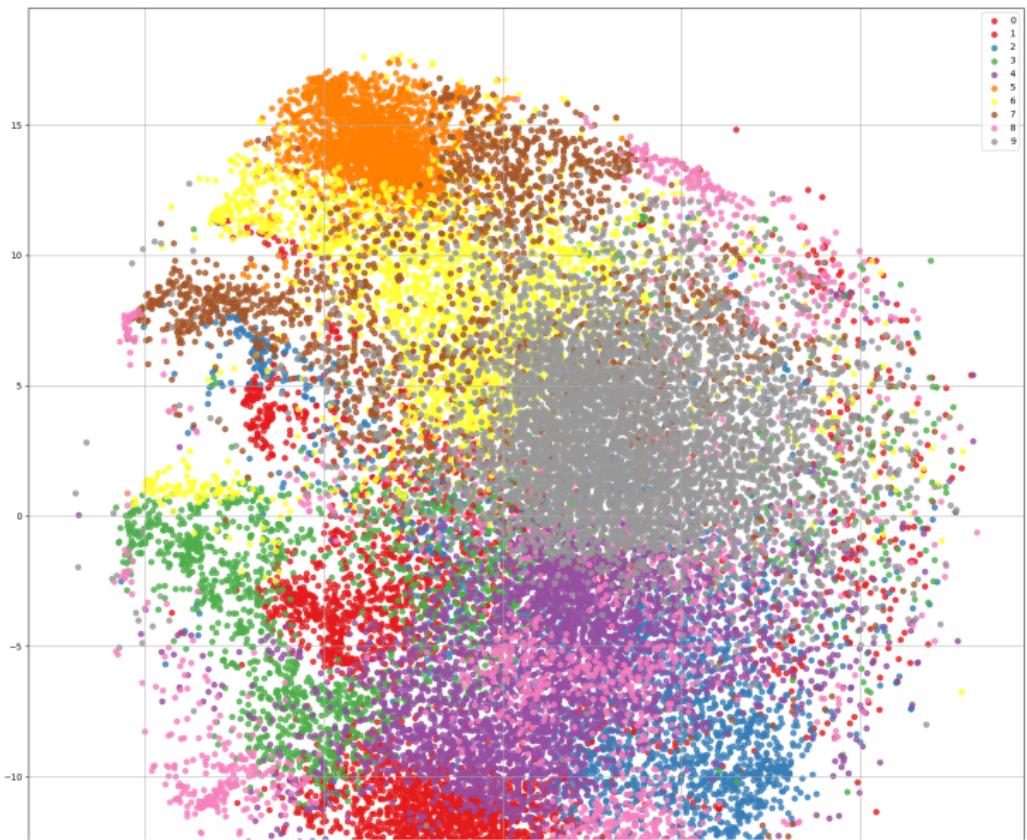
New Dataset

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

Final Remarks

- The Slightly variation between the NP grades and the high tissue variability makes difficult 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 is not the most appropriate.

Thank you...



Proposed Grading Methodology

