SEAP-IAP
[Sociedad Española de Anatomía Patológica]
[International Academy of Pathology]

INiBICA
INSTITUTO DE INVESTIGACIÓN E INNOVACIÓN BIOMÉDICA DE CÁDIZ

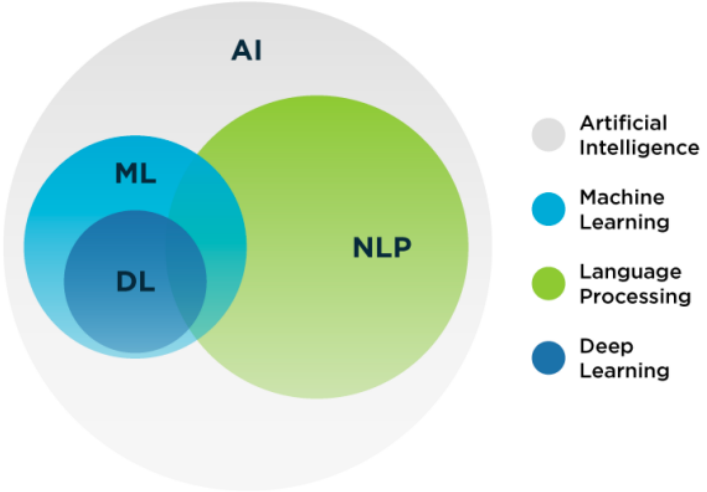
Procesamiento de Lenguaje Natural en Anatomía Patológica

Marcial García Rojo
Jefe de Servicio de Anatomía Patológica
Hospital Universitario de Jerez de la Frontera.
Cádiz. España
marcial@cim.es

9º Curso de Patología Digital. Jerez, 2022

Qué es el procesamiento de lenguaje natural (PLN, NLP)

- Es una parte de la Inteligencia Artificial
- Permite que los ordenadores (que funcionan con el llamado “lenguaje máquina”) entiendan, interpreten y manipulen el lenguaje humano



AI CAPABILITY OR METHOD

Natural Language Processing
- Chatbots & Virtual Assistants
- Machine Translation
- Voice Recognition
- Text Extraction

Machine Learning
- Supervised Learning
- Semi-supervised Learning
- Unsupervised Learning
- Classification & Clustering
- Neural Networks
- Deep Learning

- Computer Vision
- Fuzzy Logic
- Expert Systems
- Computer Heuristics
- Knowledge Bases
- Robotics
- Internet of Things (IOT)

MEDICAL SPECIALTY

CONDITION OR DISEASE STATE

DISEASE PREVENTION
- Fitness & Wellness
- Health Tracking
- Nutrition

DISEASE DIAGNOSIS
- Early Detection & Diagnosis
- Automated Image Analysis

DISEASE TREATMENT & MONITORING
- Precision Medicine
- Treatment Design
- Prognosis
- Health Monitoring
- Patient Adherence
- Robot-Assisted Surgery
- Robot-Assisted Therapy

CARE DELIVERY SUPPORT
- Clinical Workflow
- Virtual Nurses
- Care Coordination
- Dosage Error Reduction
- Medical Reference

HEALTHCARE ADMINISTRATION
- Administrative Workflow
- Customer Acquisition & Relationship Mgmt
- Clinical Insights & Risk Analysis
- Healthcare System Analysis
- Health Benefits Administration
- Data Infrastructure & Interoperability
- Patient Data Management
- Medical Education
- Cybersecurity & Fraud Detection
- Quality Measurement & Reporting

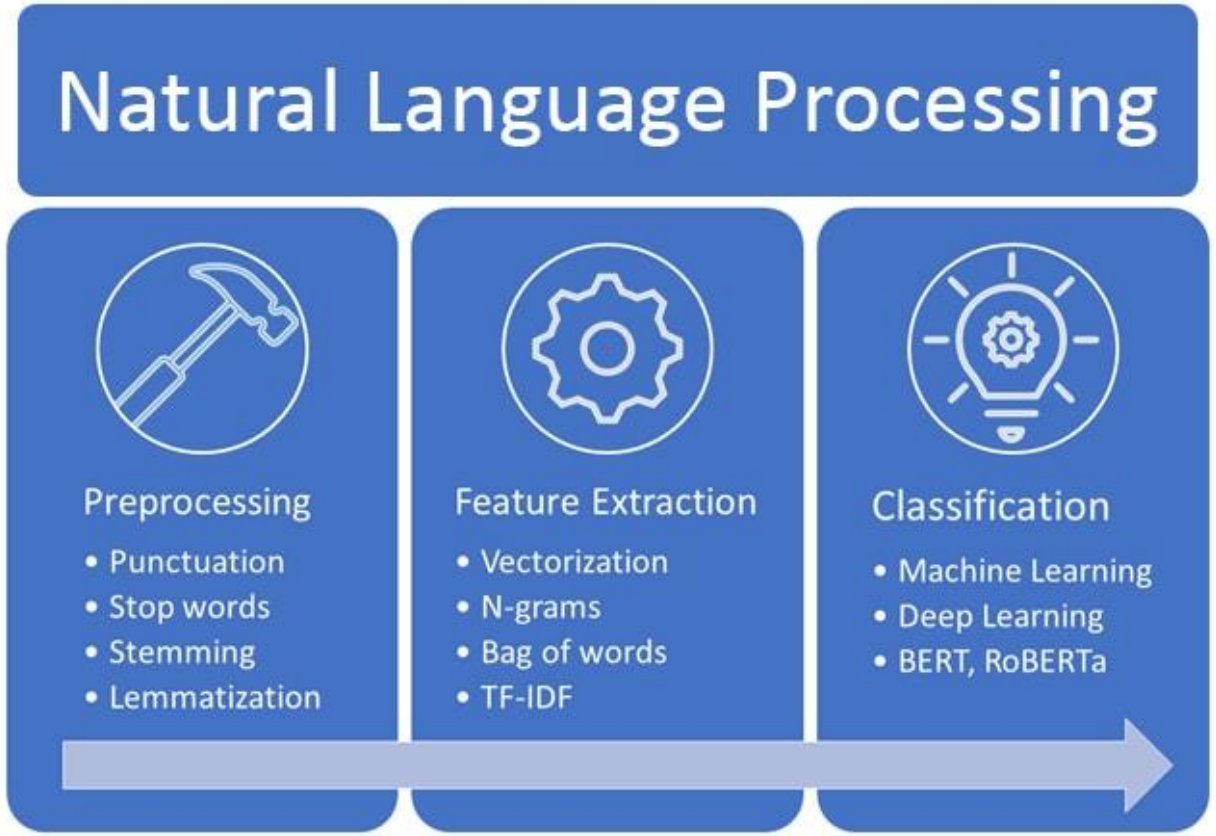
PATIENT ENGAGEMENT
- Consumer Health Information
- Digital Consultation

RESEARCH & DEVELOPMENT
- Clinical Trials
- Drug Development
- Genetic Research
- Device Development
- Diagnostic Development

From: National Academy of Medicine. Artificial Intelligence for Health Professions Educators. September 8, 2021. In: NAM.edu

Cómo funciona el PLN

- A partir de datos no estructurados (texto, ficheros de audio, mensajes), se analizan estructuras indicadas (palabras, contexto) para convertirlas en algo utilizable.
- Ejemplos: Alexa, Hey Google, chatbots…
- Incluyen algoritmos deep learning para crear unidades semánticas, eliminar palabras “innecesarias (preposiciones, artículos), buscar contextos, … y clasificar (pude combinarse con machine learning)



¿Cómo se utiliza el NLP en patología

- Para <u>clasificar</u> automáticamente informes (texto libre):

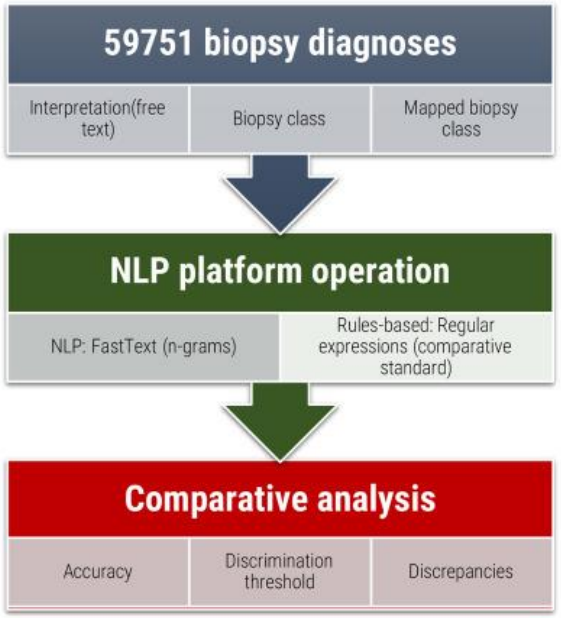
ELSEVIER

Journal of Pathology Informatics
Volume 13, 2022, 100123

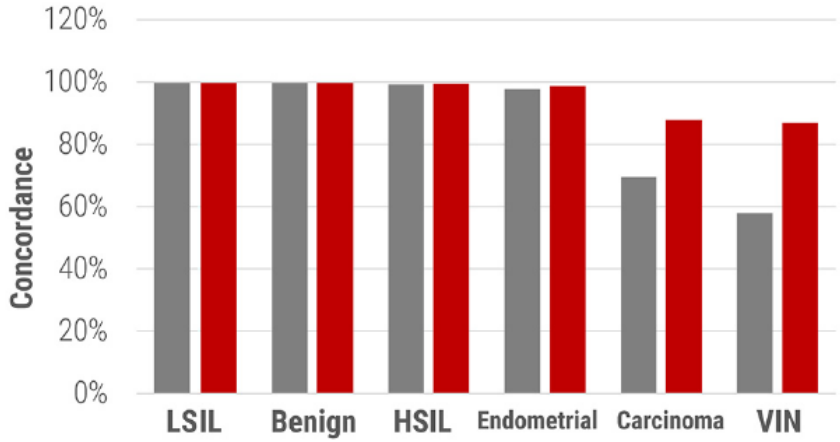
Original Research Article

Classification of cervical biopsy free-text diagnoses through linear-classifier based natural language processing

Jim Wei-Chun Hsu [a], Paul Christensen [a, b], Yimin Ge [a, b], S. Wesley Long [a, b, c]



Concordancia: 97,7%



	LSIL	Benign	HSIL	Endometrial	Carcinoma	VIN
Original	99.7%	99.7%	99.1%	97.7%	69.4%	57.9%
Optimized	99.7%	99.7%	99.4%	98.6%	87.8%	86.8%

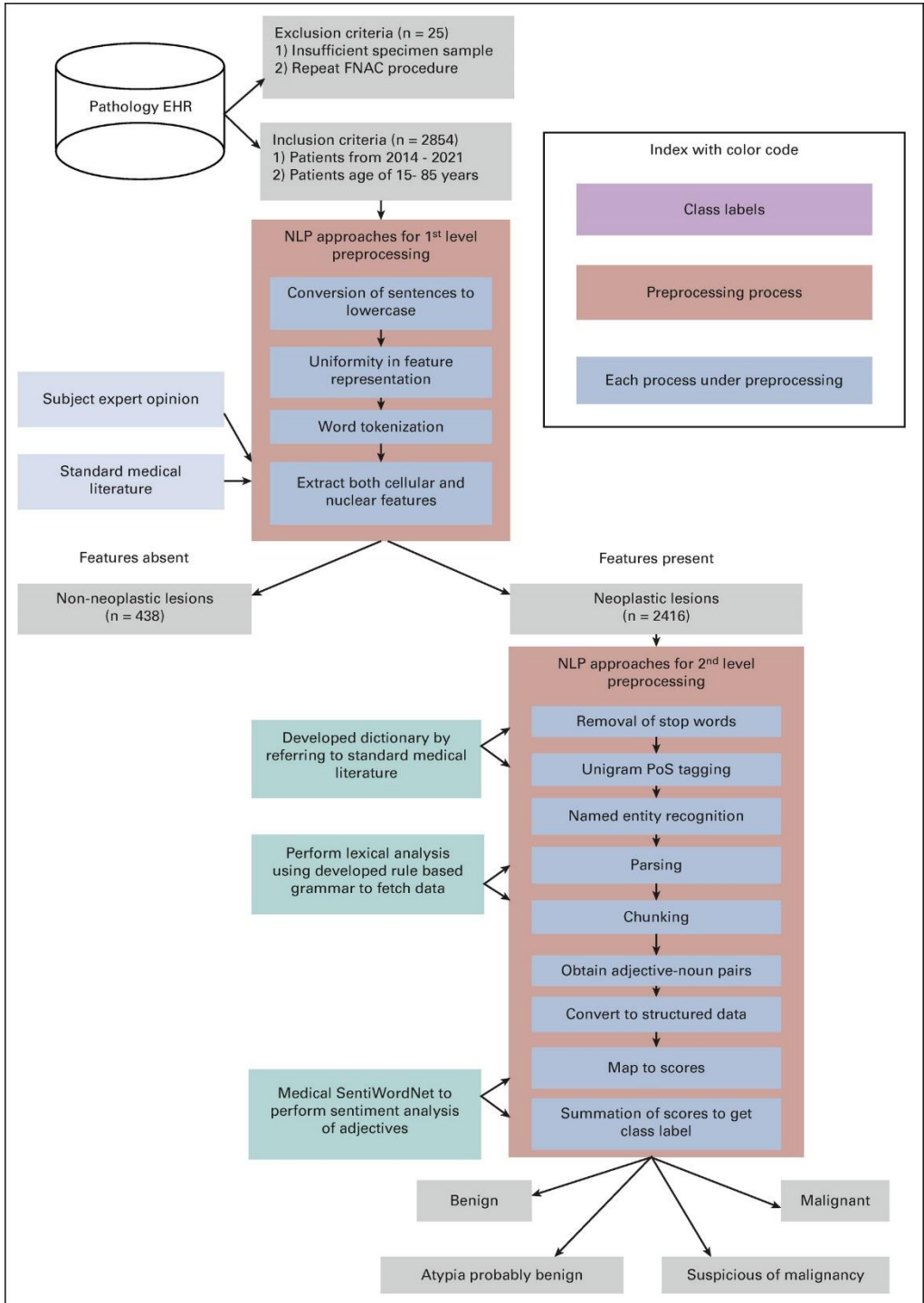
	Prediction			Referee
Interpretation	Rules	NLP	Confidence	Category
Referee concordant with Rules-based category				
scanty atypical squamous cells favor Low grade squamous intraepithelial lesion (LSIL, valN1).	VIN	LSIL	0.50	VIN
Labial condyloma (VIN1).	VIN	BENIGN	1.00	VIN
Squamous cell carcinoma in situ, at leastRare atypical glandular cells, favor neoplastic	HSIL+	CA	0.82	HSIL+
Low grade squamous intraepithelial lesion (CIN 1) with microfocal moderate dysplasia/CIN 2	HSIL+	LSIL	0.87	HSIL+
Invasive squamous cell carcinoma, well differentiated, arising in a background of high grade squamous intraepithelial lesion (HSIL, CIN 3).Tumor invades at least 1.5 mm deep.Deep / endocervical margin is positive for invasive carcinoma.Endocervical and ectocervical surgical margins are positive for HSIL.	CA	HSIL+	0.96	CA
Referee concordant with NLP-based category				
Features consistent with exogenous progestin effect	BENIGN	ENDOMETRIUM	0.75	ENDOMETRIUM
Minute fragment of squamous epithelium with mild atypia (see microscopic description)Fragments of superficial endometrial tissue (see microscopic description)	ENDOMETRIUM	BENIGN	0.30	BENIGN
Scant transformation zone mucosa with squamous metaplasia, inflammation, and reactive changesNegative for dysplasiaBenign endometrial tissue, proliferative pattern	ENDOMETRIUM	BENIGN	1.00	BENIGN
Low grade squamous intraepithelial lesion (LSIL, CIN 1) Endocervical mucosa/transformation zone not present.	BENIGN	LSIL	0.73	LSIL
Minute focus of High Grade Squamous Intraepithelial lesion Confirmed on P16/ki67 block positivitybenign endocervical mucosa	BENIGN	HSIL+	0.96	HSIL+
Referee discrepant with both categories				
Stratified mucin-producing intraepithelial lesion (SMILe; see comment)	BENIGN	CA	0.34	HSIL+
polypoid fragment of vulvar skin with focal features compatible with human papillomavirus infection/ condyloma accuminatum. see note.note: Immunostain for Ki-67 (a proliferation marker) has been reviewed. A few suprabasilar keratinocytes show positive nuclear staining with Ki-67. This staining pattern supports the above diagnosis.	BENIGN	LSIL	0.99	VIN

Nandish S et al. JCO Clinical Cancer Informatics no. 6 (2022) e2200036.

Natural Language Processing Approaches for Automated Multilevel and Multiclass Classification of Breast Lesions on Free-Text Cytopathology Reports

RESULTS

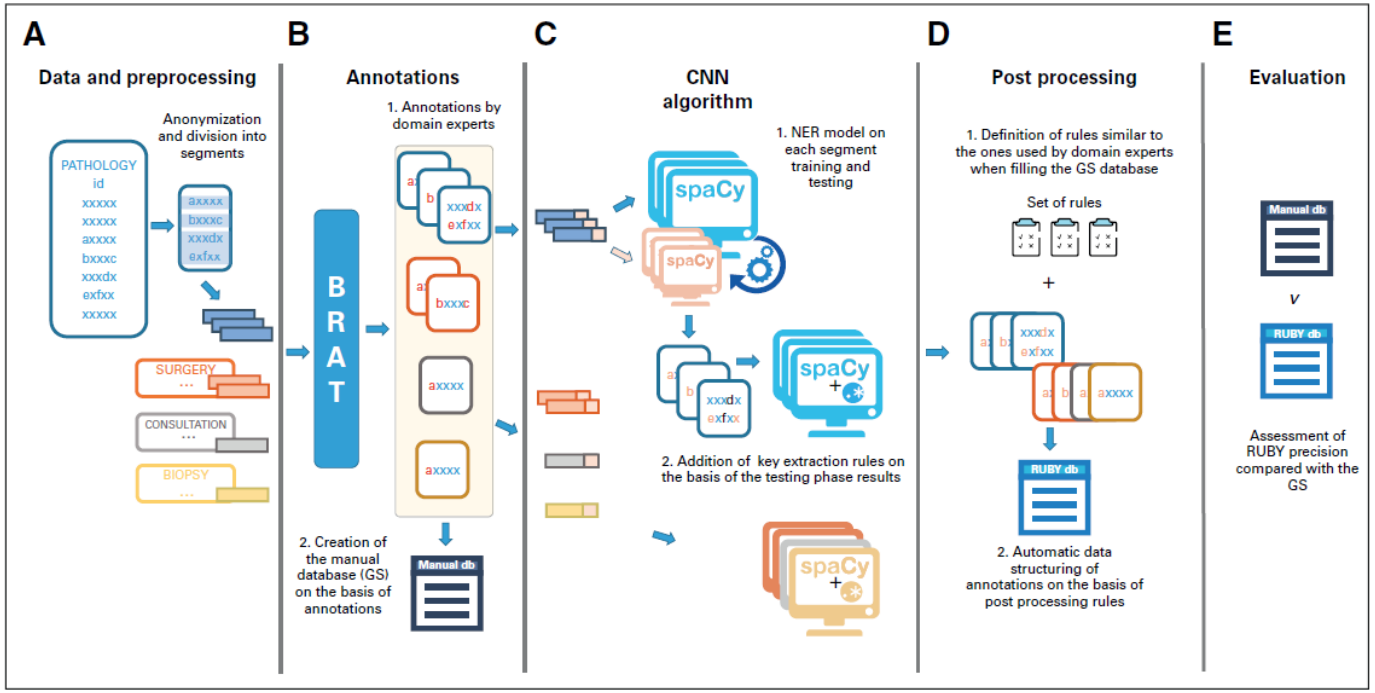
Random forest and feed-forward neural network classifiers gave the best performance with an accuracy of 99.36%, an overall receiver operating characteristic-area under the curve of 99.2%, a correlation with ground truth of 98.3%, and a histopathologic correlation of 98.6%.

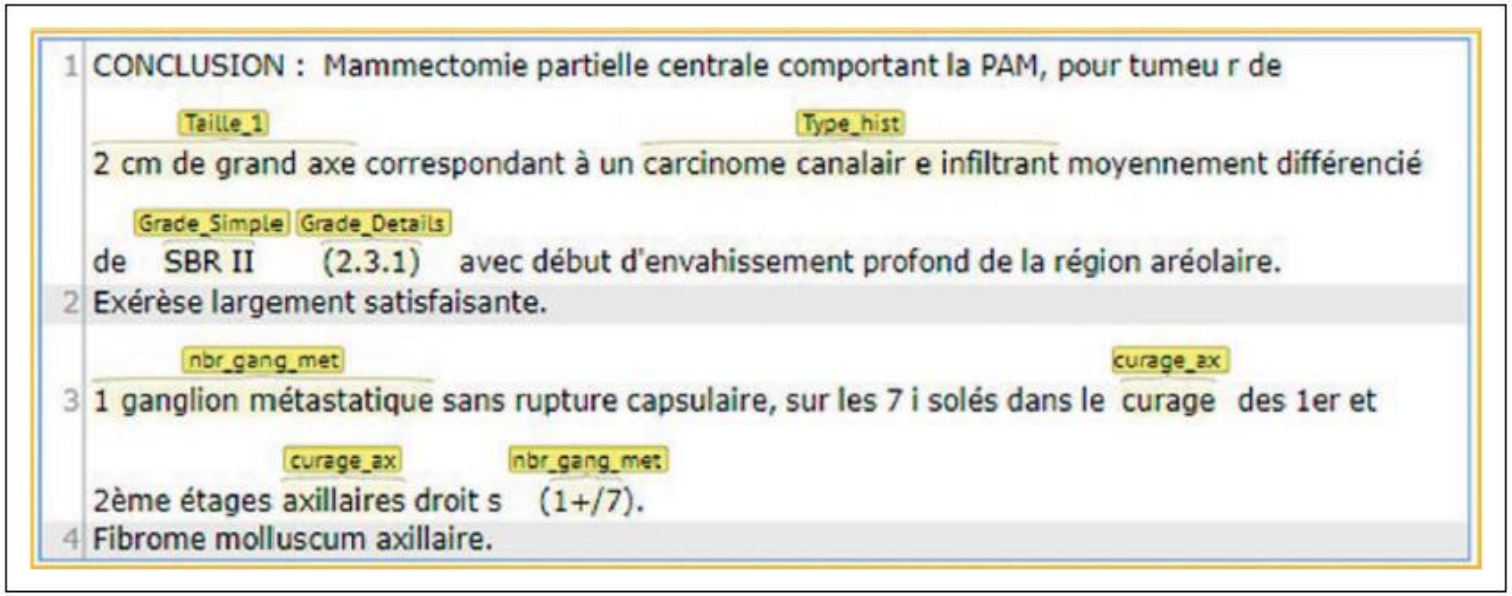


Para seleccionar informes

- Wenker et al (14 sep 2022), Clin Gastroenterol Hepatol: Identificar displasia en informes de esófago de Barrett: 600 pacientes (60 con displasia) y 400 para validación (39 con displasia) → 98,7% exactitud, 100% precisión.
- Gray et al (6 sep 2022), JCO Clin Cancer Inform: Identificar pacientes de cáncer de ovario con test BRCA1/2: 0,9979 precisión, F-score: 0,9977.
- Schiappa et al (jul 2022) JCO Clin Cancer Inform: En francés, cáncer de mama: precisión de 92,8% en piezas, 98,1% en biopsias y 81,8% en consultas.
- Legnar et al. Diagnostics 2022, 12, 1726. En alemán, nefropatología, clasificar en grupos diagnósticos. Máximo F1-scores : 0,892 (insuficiente en algunas entidades)

RUBY: NLP of French EMRs for Breast Cancer Research





Schiappa et al (jul 2022)
JCO Clin Cancer Inform

Report	Segment	Entity	No.	Precision	Recall	F1 Score
Surgery	Title	Type of surgery	544	93.3	92.8	93.0
		Search for SLN	279	97.1	98.9	98.0
Pathology	General Conclusion	Histological type	67	76.1	63.6	69.3
		Lesion size 1, mm	48	85.7	88.2	87.0
		SBR grade (Elston and Ellis)	46	93.8	88.2	90.9
	HER2	HER2 result (IHC)	32	93.8	97.0	95.2
	Hormone Receptors	Estrogen expression (positive)	54	87.9	96.7	92.1
		Progesterone expression (positive)	36	88.5	88.5	88.5
Biopsy	Biopsy	Histological type	76	83.6	83.6	83.6
		HER2 realization	28	96.3	96.3	96.3
		HER2 result (IHC)	18	100.0	100.0	100.0

Schiappa et al (jul 2022) JCO Clin Cancer Inform

words that strongly indicate a cluster name are highlighted in green (strong topic words). Orange highlighted topic words only weakly indicate a cluster name (weak topic words).

Cluster Index-Cluster Name	Keywords According to SVM
0-systemic lupus erythematosus	scale, chronicity_index, activity_index, class, -nih, iv, lupus_nephritis, component, iii, who
1-rapid progressive glomerulonephritis	quantity, sclerosing, glomeruli, scarred, glomerulus, fresh, of_which1, proliferating_glomerulonephritis, necrosis, sclerosed
2-tubulo-interstitial nephritis	tubular_epithelial_damage, moderate, minor, damage, none, completion, fibrosis, finally, known, tubulointerstitial
3-pauci immune glomerulonephritis	of_this, total_amount, intact, pauci_immune_glomerulonephritis, necrotizing, glomeruli, scarred, *, remaining, extracapillary
4-IgA nephritis	oxford_classification, e0, s1, m1, t0, c0, iga_glomerulonephritis, s0, applicable, e1
5-FSGN	fsg, primary, distinction, look_together, secondary, segmental_glomerulosclerosis, collapsing, at_most, continuing, patient
6-thrombotic microangiopathy	microangiopathy, thrombotic, reparation, preglomerular, glomerular, located, active, glomerular_thrombotic, hypertension, overwhelmingly
7-multiple myeloma	cast_nephropathy, myeloma_kidney, lcdd, lambda, kappa, amyloidosis, followed_by, light_chains, al-amyloidosis, light_chain_nephropathy
8-kidney transplant	rejection, routinely, success, examination, calcineurin_inhibitor_toxicity, ascending, humorous, bacterial, urinary_tract_infection, follow-up_report

the clusters 1 (systemic lupus erythematosus), 2 (tubulo-interstitial nephritis), and 3 (pauci immune glomerulonephritis) could be recognized well by all classifiers

Legnar et al. Diagnostics 2022, 12, 1726

Extraer información relevante

- Enlazar preparaciones digitales con informes para evitar tener que anotar las preparaciones una a una.

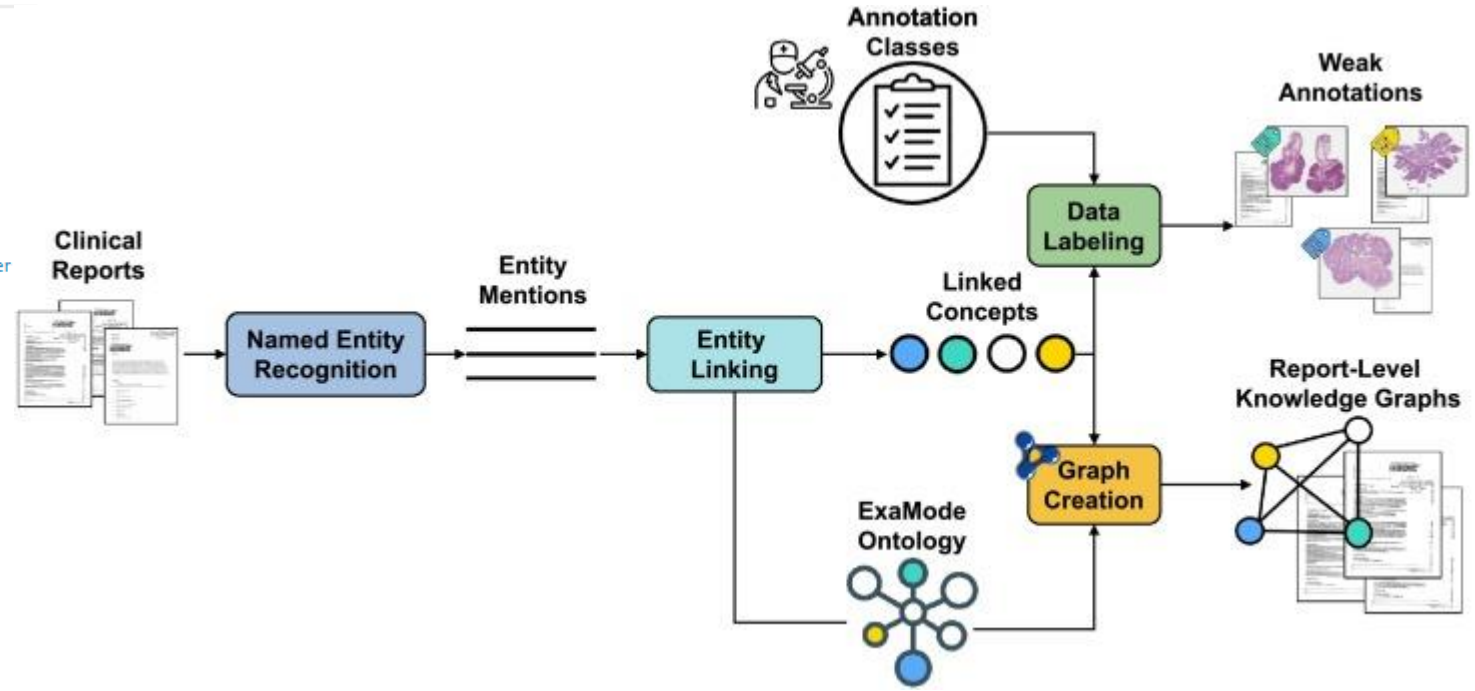
ELSEVIER

Journal of Pathology Informatics
Volume 13, 2022, 100139

Original Research Article

Empowering digital pathology applications through explainable knowledge extraction tools

Stefano Marchesin [a], Fabio Giachelle [a], Niccolò Marini [b], Manfredo Atzori [b, c], Svetla Boytcheva [d], Genziana Buttafuoco [e], Francesco Ciompi [f], Giorgio Maria Di Nunzio [a], Filippo Fraggetta [e], Ornella Irrera [a], Henning Müller Todor Primov [d], Simona Vatrano [e], Gianmaria Silvello [a]



Revisión de NLP en patología

The American Journal of Pathology, Vol. ■, No. ■, ■ 2022

ELSEVIER

The American Journal of
PATHOLOGY

ajp.amjpathol.org

REVIEW

Natural Language Processing in Pathology

Current Trends and Future Insights

Pilar López-Úbeda,* Teodoro Martín-Noguerol,† José Aneiros-Fernández,‡ and Antonio Luna†

From the R+D+I Department and the MRI Unit,† Radiology Department, HT Medica, Jaén; and the Department of Pathology,‡ HT Medica, Granada, Spain*

PROOF

Conclusiones

- Además del reconocimiento de voz, hay muchas aplicaciones del PLN que son de interés en anatomía patológica.
- Para la clasificación y selección de informes en áreas de oncología ya funciona muy bien.
- En algunas áreas (nefropatología) es necesario mejorar los algoritmos disponibles. ¿Bastaría aumentar nº de casos? ¿mejorar algoritmos combinando deep learning y machine learning?
- PLN puede ayudar para anotar preparaciones digitales y obtener grandes repositorios de imágenes “anotadas”

¡Gracias!

Marcial Garcia Rojo
marcial@cim.es