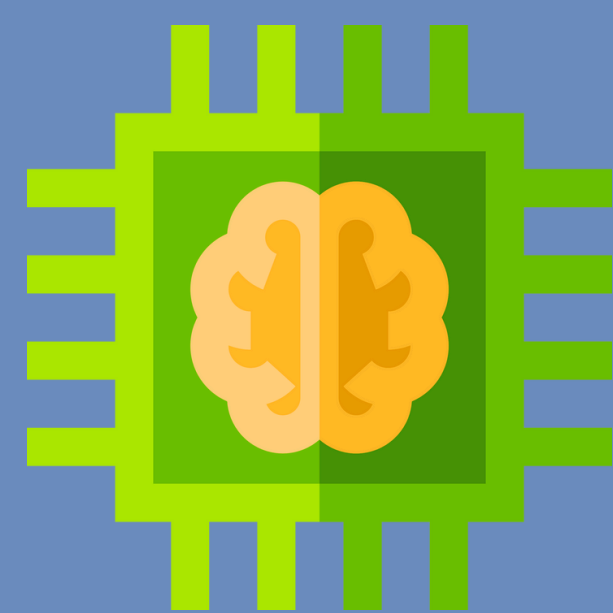


Artificial Intelligence transforms medical text into structured data ready for research



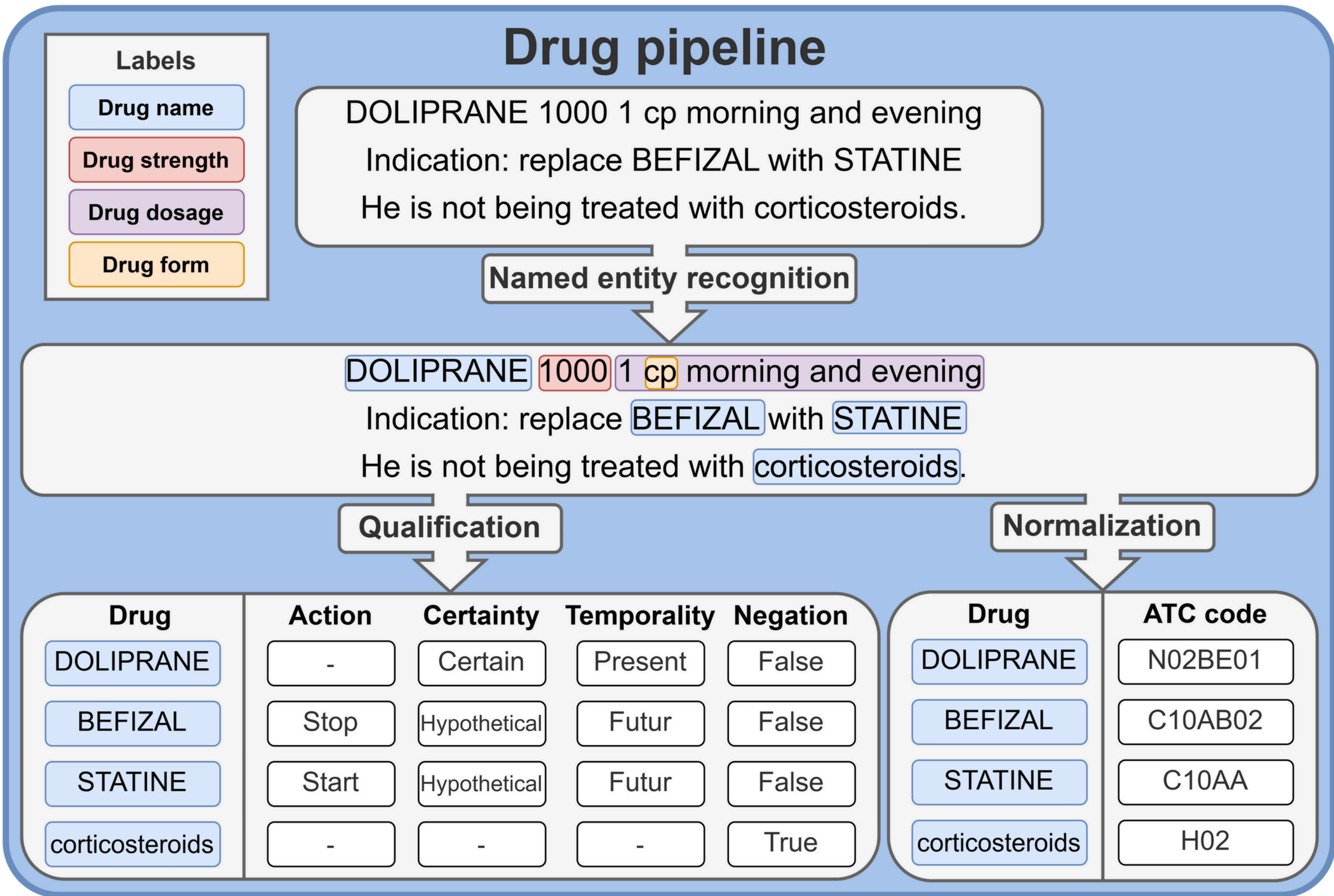
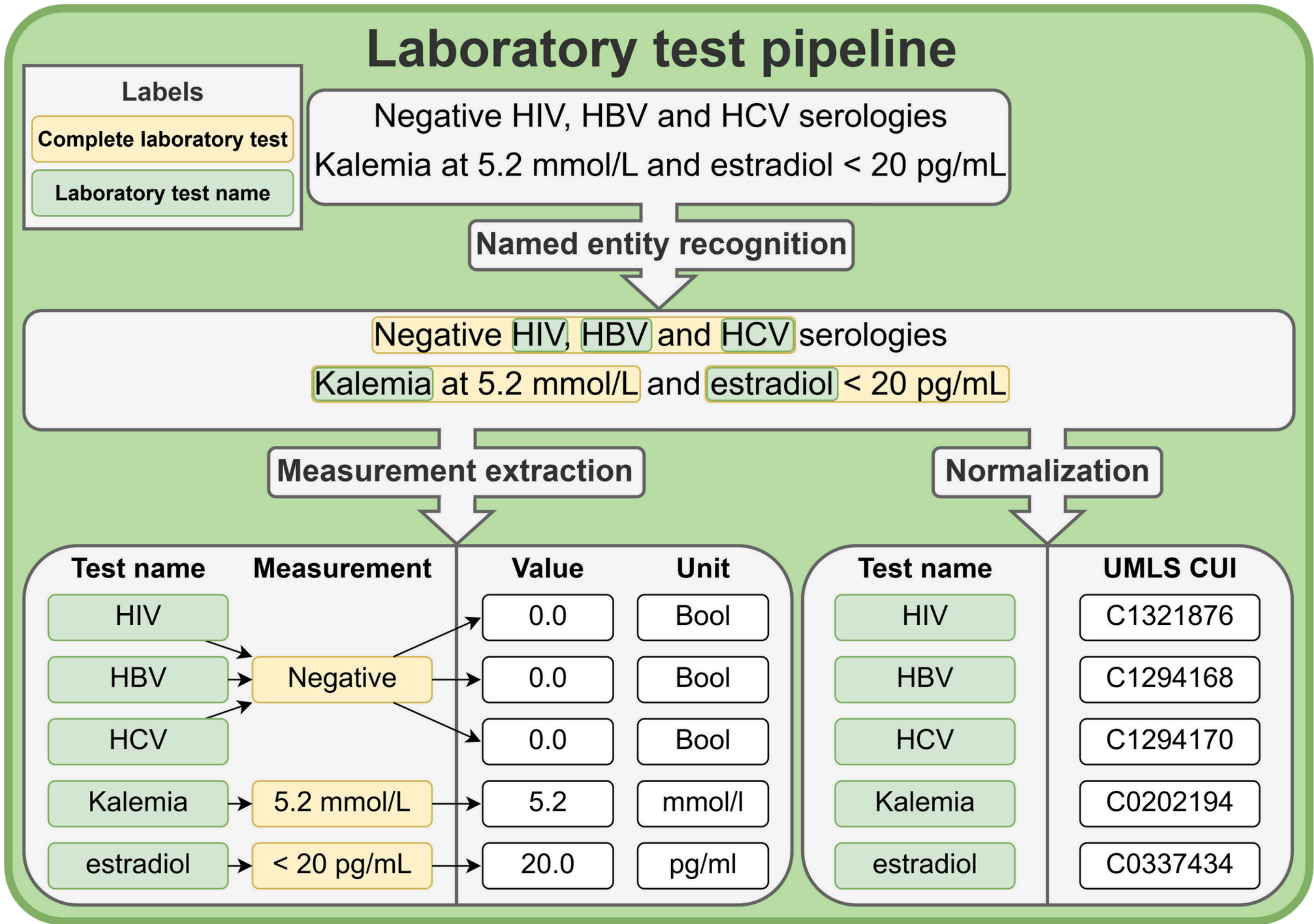
Unstructured data



Structured data

Background: Healthcare digitization through electronic health records (EHRs) creates new valuable research opportunities. While EHRs consist of both structured and unstructured data, unstructured data, such as reports, contain critical information **missing from structured data**.

Method: This transformation involves 4 **Natural Language Processing** tasks: Named Entity Recognition, Qualification, Measurement Extraction and Normalization.



Validation of the pipelines: They are trained and evaluated on **103 annotated medical texts**.

Label	Number of entities	Precision	Recall	F1-score
Laboratory test	356 [204-548]	72.0 [65.3-79.1]	70.2 [59.7-77.9]	71.1 [63.6-77.8]
Drug name	312 [206-424]	91.9 [88.9-93.5]	86.9 [82.9-90.0]	89.3 [85.9-91.6]

Application of the pipelines: They are applied to a cohort of 4102 patients diagnosed with Systemic Lupus Erythematosus, corresponding to **12500 medical reports** and 10445 EHRs.

Laboratory tests

Laboratory test	No. of patients with positive tests (ratio)		
	Structured data only	Structured and unstructured data	Benefits of the unstructured data
Antinuclear antibody	752 (18.3%)	2949 (71.9 %)	+ 2197 (53.6%)
Anti-DNA antibodies	541 (13.2%)	2174 (53.0 %)	+ 1633 (39.8%)
Anti-Sm	255 (6.2%)	858 (20.9 %)	+ 603 (14.7 %)

Drug treatments

Drug	No. of patients with drug treatments (ratio)		
	Structured data only	Structured and unstructured data	Benefits of the unstructured data
Glucocorticoids	950 (23.2%)	3308 (80.6%)	+ 2358 (57.5%)
Cyclophosphamide	64 (1.6%)	894 (21.8%)	+ 830 (20.2%)
Mycophenolate	301 (7.3%)	1263 (30.8%)	+ 962 (23.5%)
Rituximab	75 (1.8%)	709 (17.3%)	+ 634 (15.5%)
Belimumab	43 (1.0%)	247 (6.0%)	+ 204 (5.0%)
Methotrexate	112 (2.7%)	963 (23.5%)	+ 851 (20.7%)
Hydroxychloroquine	920 (22.4%)	3520 (85.8%)	+ 2600 (63.4%)



Paper



GitHub



Contact

Authors: Adam Remaki, Jacques Ung, Pierre Pages, Perceval Wajsbürt, Guillaume Faure, Thomas Petit-Jean, Xavier Tannier, Christel Gérardin

