

# Natural Language Processing to extract SNOMED-CT codes from pathological reports

Giorgio Cazzaniga<sup>1</sup>, Albino Eccher<sup>2</sup>, Enrico Munari<sup>3</sup>, Stefano Marletta<sup>3</sup>, Emanuela Bonoldi<sup>4</sup>, Vincenzo Della Mea<sup>5</sup>, Moris Cadei<sup>6</sup>, Marta Sbaraglia<sup>7</sup>, Angela Guerriero<sup>7</sup>, Angelo Paolo Dei Tos<sup>7</sup>, Fabio Pagni<sup>1</sup>, Vincenzo L'Imperio<sup>1</sup>

<sup>1</sup> Department of Medicine and Surgery, Pathology, IRCCS Fondazione San Gerardo dei Tintori, University of Milano-Bicocca, Italy; <sup>2</sup> Section of Pathology, Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, University Hospital of Modena, Modena, Italy; <sup>3</sup> Department of Diagnostic and Public Health, Section of Pathology, University of Verona, Verona, Italy; <sup>4</sup> Unit of Surgical Pathology and Cytogenetics, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy; <sup>5</sup> Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy; <sup>6</sup> Pathology Unit, ASST Spedali Civili di Brescia, Brescia, Italy; <sup>7</sup> Surgical Pathology and Cytopathology Unit, Department of Medicine-DIMED, University of Padua School of Medicine, Padua, Italy

## Summary

**Objective.** The use of standardized structured reports (SSR) and suitable terminologies like SNOMED-CT can enhance data retrieval and analysis, fostering large-scale studies and collaboration. However, the still large prevalence of narrative reports in our laboratories warrants alternative and automated labeling approaches. In this project, natural language processing (NLP) methods were used to associate SNOMED-CT codes to structured and unstructured reports from an Italian Digital Pathology Department.

**Methods.** Two NLP-based automatic coding systems (support vector machine, SVM, and long-short term memory, LSTM) were trained and applied to a series of narrative reports.

**Results.** The 1163 cases were tested with both algorithms, showing good performances in terms of accuracy, precision, recall, and F1 score, with SVM showing slightly better performances as compared to LSTM (0.84, 0.87, 0.83, 0.82 vs 0.83, 0.85, 0.83, 0.82, respectively). The integration of an explainability allowed identification of terms and groups of words of importance, enabling fine-tuning, balancing semantic meaning and model performance.

**Conclusions.** AI tools allow the automatic SNOMED-CT labeling of the pathology archives, providing a retrospective fix to the large lack of organization of narrative reports.

**Key words:** SNOMED-CT, digital pathology, natural language processing, laboratory information system

## Introduction

The progressive digitization of pathology laboratories represents a valuable opportunity to promote automation and standardization of the diagnostic process. This can be perceived through the progressive introduction of artificial intelligence (AI) algorithms for the most disparate tasks <sup>1</sup>, even in highly sub-specialized fields <sup>2-5</sup>, showcasing the incredible potentialities of these new technologies in reducing interobserver variability and optimizing time-consuming tasks. On the other hand, the modifications of the pathology workflow towards a fully tracked and controlled system through the implementation of innovative instruments <sup>6,7</sup> and increasing attention to the sample archives <sup>8,9</sup> is significantly twisting our routine practice, leading international societies to promote guidelines and recommendations to guide in this intricate transition <sup>10</sup>. The

Received November 16, 2023  
Accepted: November 17, 2023

### Correspondence

Vincenzo L'Imperio  
Department of Medicine and Surgery, University of Milano-Bicocca  
IRCCS Fondazione San Gerardo dei Tintori,  
via Pergolesi 33, 20900, Monza (Italy)  
E-mail: vincenzo.limperio@gmail.com

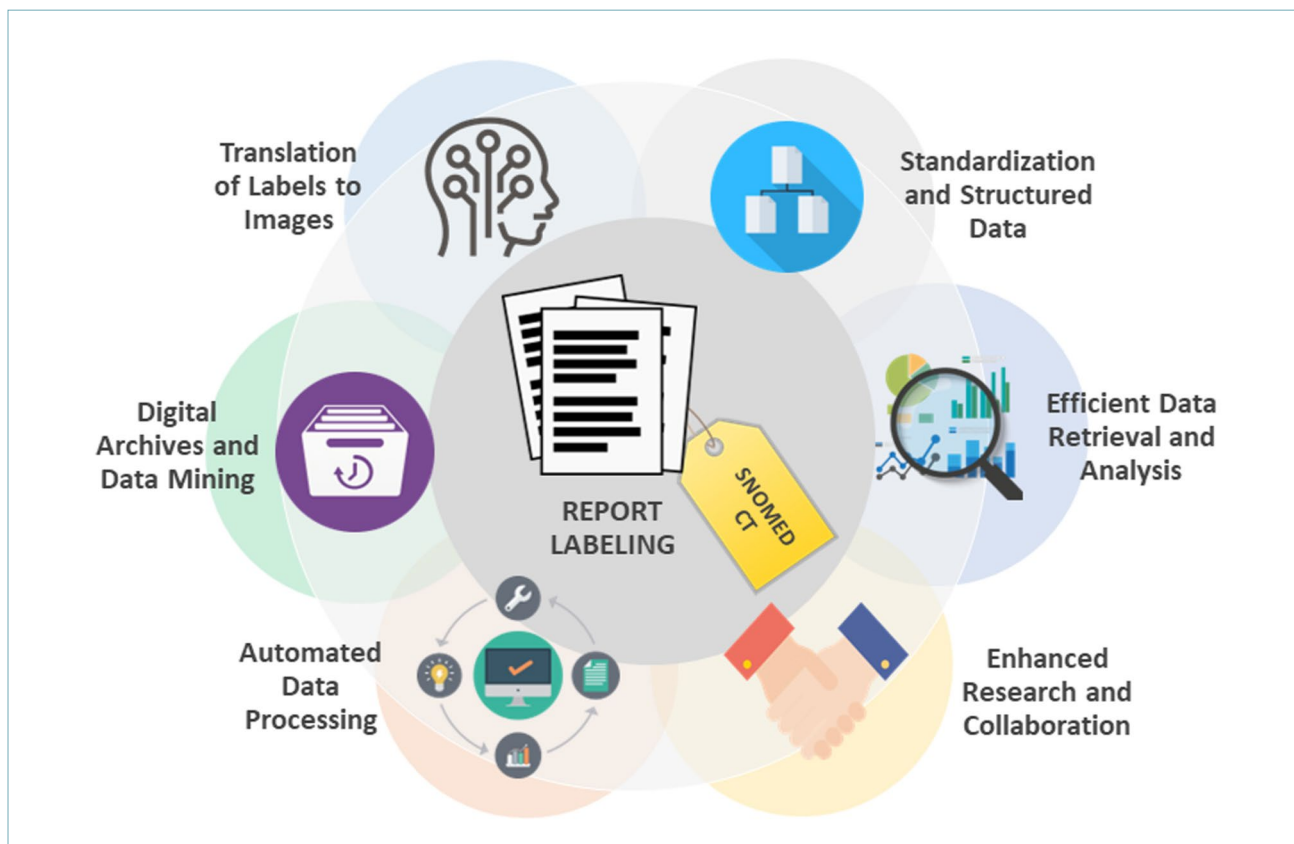
**How to cite this article:** Cazzaniga G, Eccher A, Munari E, et al. Natural Language Processing to extract SNOMED-CT codes from pathological reports. *Pathologica* 2023;115:318-324. <https://doi.org/10.32074/1591-951X-952>

© Copyright by Società Italiana di Anatomia Patologica e Citopatologia Diagnostica, Divisione Italiana della International Academy of Pathology



OPEN ACCESS

This is an open access journal distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license: the work can be used by mentioning the author and the license, but only for non-commercial purposes and only in the original version. For further information: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



**Figure 1.** The multiple benefits of labeling diagnosis. **Standardization and Structured Data:** labeled diagnoses can be consistently categorized, making it easier to analyze and compare data across different cases and institutions. **Data Retrieval and Analysis:** with structured data, specific diagnoses can be quickly searched and extracted, allowing pathologists and researchers to access relevant cases more easily. **Enhanced Research and Collaboration:** researchers can perform retrospective studies by extracting cases with specific diagnosis from digital archives, enabling large-scale analyses and exploration of various research questions. **Automated Data Processing:** by having structured labels, it becomes possible to develop algorithms and machine learning models that can automatically process pathology data. **Digital Archives and Data Mining:** digital archives can be organized and indexed based on diagnoses, enabling efficient retrieval and utilization of archived cases. **Translation of Labels to Images:** labeled diagnoses can be translated to images, opening avenues for weakly supervised machine learning projects, even without explicit pixel-level annotations.

introduction of standardized structured reports (SSR) represents another milestone of the digital pathology transition in the precision medicine era, either full<sup>11</sup> or partial<sup>12,13,14</sup>. Waiting for the progressive adoption of SSR in our routine practice, the application of natural language processing (NLP) methods for the automated labeling of retrospective series with Systematized Nomenclature of Medicine Clinical Term (SNOMED-CT) is a promising alternative to untap the goldmine of the digital pathology archives (Fig. 1). The retrieval of this information holds the potential for conducting high-quality analyses and research, aiming to enhance standards of care and diagnosis, ultimately leading to improved patient outcomes<sup>15</sup>, being poten-

tially integrable within our laboratory information systems (LIS)<sup>16</sup>.

The identification of robust supports for data mining and curation may be the turning point for building *data lakes* projects in pathology, as a test possibly expandable to many other hospital departments. Here we investigate whether an automated NLP approach for the SNOMED-CT code labeling on a prospective cohort of structured and unstructured reports from an Italian Digital Pathology Department with two different AI algorithms can optimize the processing of highly specialized languages by developing a scalable, lightweight, and efficient model.

## Materials and methods

### DATA EXTRACTION AND FILTERING

Narrative and structured reports, along with their corresponding SNOMED-3 codes, were extracted from the LIS Athena (Version 4.3.0, Dedalus, Florence, Italy) at the Pathology Department of Fondazione IRCCS San Gerardo dei Tintori, University of Milano-Bicocca, Monza, Italy for cases diagnosed from January 2023 to November 2023. The dataset underwent a series of filtering steps aimed to clean it from cases lacking a SNOMED code and those containing multiple lesions within the same diagnosis text resulting in multiple coding. Specifically, only diagnosis (D) and morphology (M) codes were retained and appropriately translated to SNOMED-CT codes.

### AD HOC ONTOLOGY

Ad hoc ontology was constructed based on the International Classification of Diseases SNOMED-CT. In instances where a report contained multiple morphological codes, we selected the one that indicated more advanced disease progression or held greater clinical significance as the primary representation. Conversely, cases where multiple lesions with equivalent potential for malignancy or not linked by a single pathogenetic process were excluded from the analysis.

### PREPROCESSING

The dataset underwent a series of preprocessing steps to prepare it for testing on the models trained as demonstrated in the following repository: <https://github.com/Gizmopath/SVM-LSTM-models-for-automated-SNOMED-CT-labeling>. This process included retaining the 50 most frequent codes along with their corresponding diagnoses. Outliers, such as unusually lengthy diagnosis texts exceeding 750 characters, were removed. Additionally, class balancing techniques were implemented to address potential biases in the data, ensuring equal representation of different classes. Stopwords, special characters, and punctuation marks were removed and text was converted into lowercase. Italian lemmatization and tokenization processes were also performed.

### NATURAL LANGUAGE PROCESSING MODELS PERFORMANCES

Two different models, long short-term memory (LSTM) and support vector machine (SVM), were tested for their ability to assign the correct SNOMED-CT code to the relative diagnostic report. The models' performances were evaluated keeping track of accuracy, precision, recall and F1 score as metrics, providing insights into their predictive capabilities.

### EXPLAINABILITY

Explainability was incorporated into the models using the Local Interpretable Model-agnostic Explanations (LIME) library. This approach allowed the identification of the most relevant terms associated with each SNOMED-CT code category. Before the analysis with LIME library, words such as "eventual", "useful", "evaluation", "confirm", "appears", etc. were therefore gradually eliminated during the preprocessing phase to achieve a smoother workflow and avoid classification errors stemming from words that are semantically important but not medically/diagnostically relevant. By examining the explanations provided by LIME, insights were gained into the features and patterns that influenced the model's predictions.

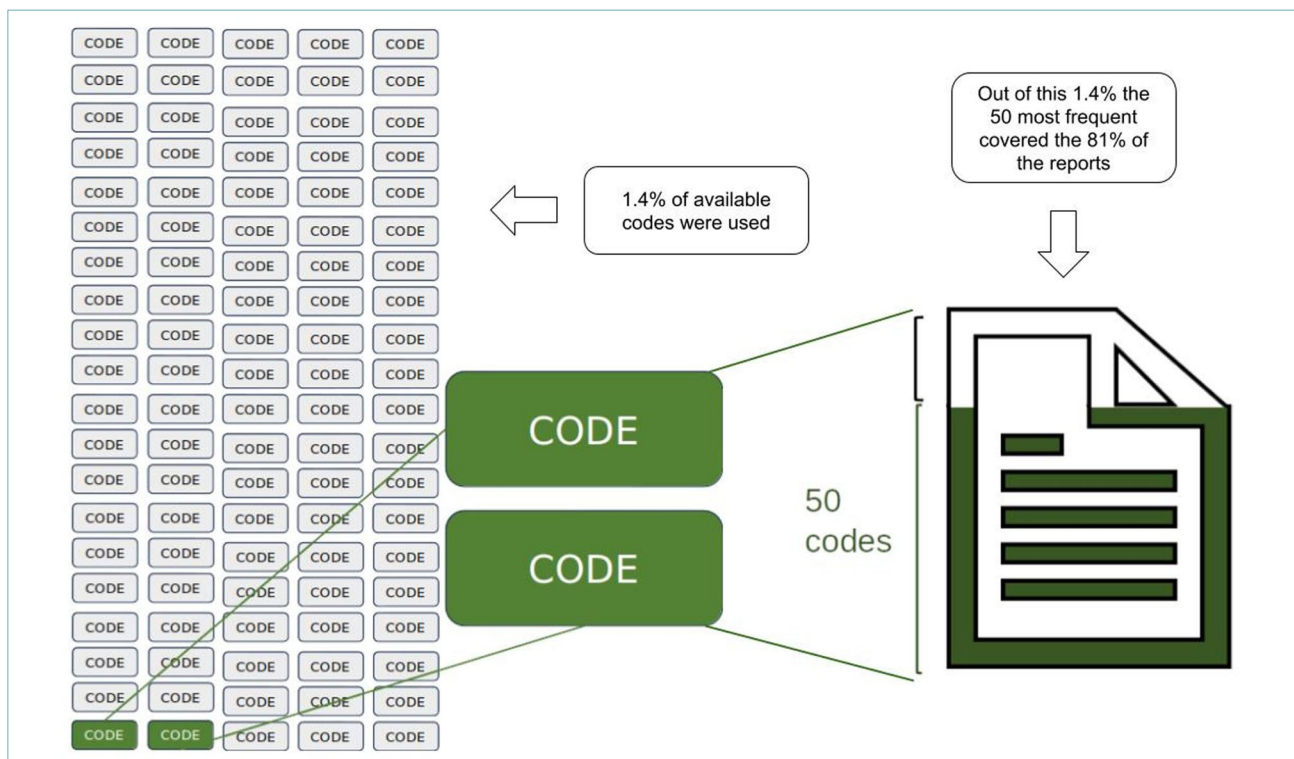
## Results

### CASES

A total of 1163 reports from 2023 were retrieved and used to test the two different NLP algorithms. The translation of SNOMED-3 codes to SNOMED CT resulted in a total of 46 'Morphologic abnormality' codes, 1 'Disorder' code, 2 'Finding' codes and 1 'Quantitative' code. The most frequent codes were associated with dermatopathology and gastroenterology, as well as more generic codes such as "Normal tissue morphology" or "chronic inflammation", reflecting the case frequencies in the Department (Supplementary Tab. I). The restriction of the dataset to the most 50 used codes allowed us to work on codes where we had a substantial sample size still being able to cover 81% of the initial dataset. The choice to narrow down the dataset to the top 50 cases was made to strike a balance between maintaining a high dataset coverage percentage and ensuring consistent accuracy results. Expanding the dataset to 70 or 100 cases would have increased the coverage from 81% to 86% and just under 90%, respectively (Fig. 2).

### NLP MODELS PERFORMANCES

The comparison of the performances obtained with LSTM and SVM models on the prospective 2023 cohort test set is reported in Table I. Although not reaching statistical significance, the SVM model showed similar or even better results as compared to the LSTM one in terms of accuracy, precision, recall and F1 score. The code assigned with worst performance was M-09410 (in SNOMED-CT: 'Negative for residual tumor (Finding)', F1 score: 0.26). This is related to a common difficulty in NLP when dealing with negations, which needs to be treated with specific techniques.



**Figure 2.** The restriction of the dataset to the most 50 used codes allowed us to work on codes where we had a substantial sample size still being able to cover 81% of the initial dataset. The original number of unique codes in our dataset was 1251, representing about 1.4% of the sum of all morphological abnormality (M-) and disease (D-) SNOMED-3 codes.

**EXPLAINABILITY AI FOR SNOMED-KEYWORDS CORRESPONDENCE**

The application of explainability algorithms for the detection of human interpretable features within the narrative report certified the genuinity of the SNOMED-CT prediction by the NLP approach, highlighting the

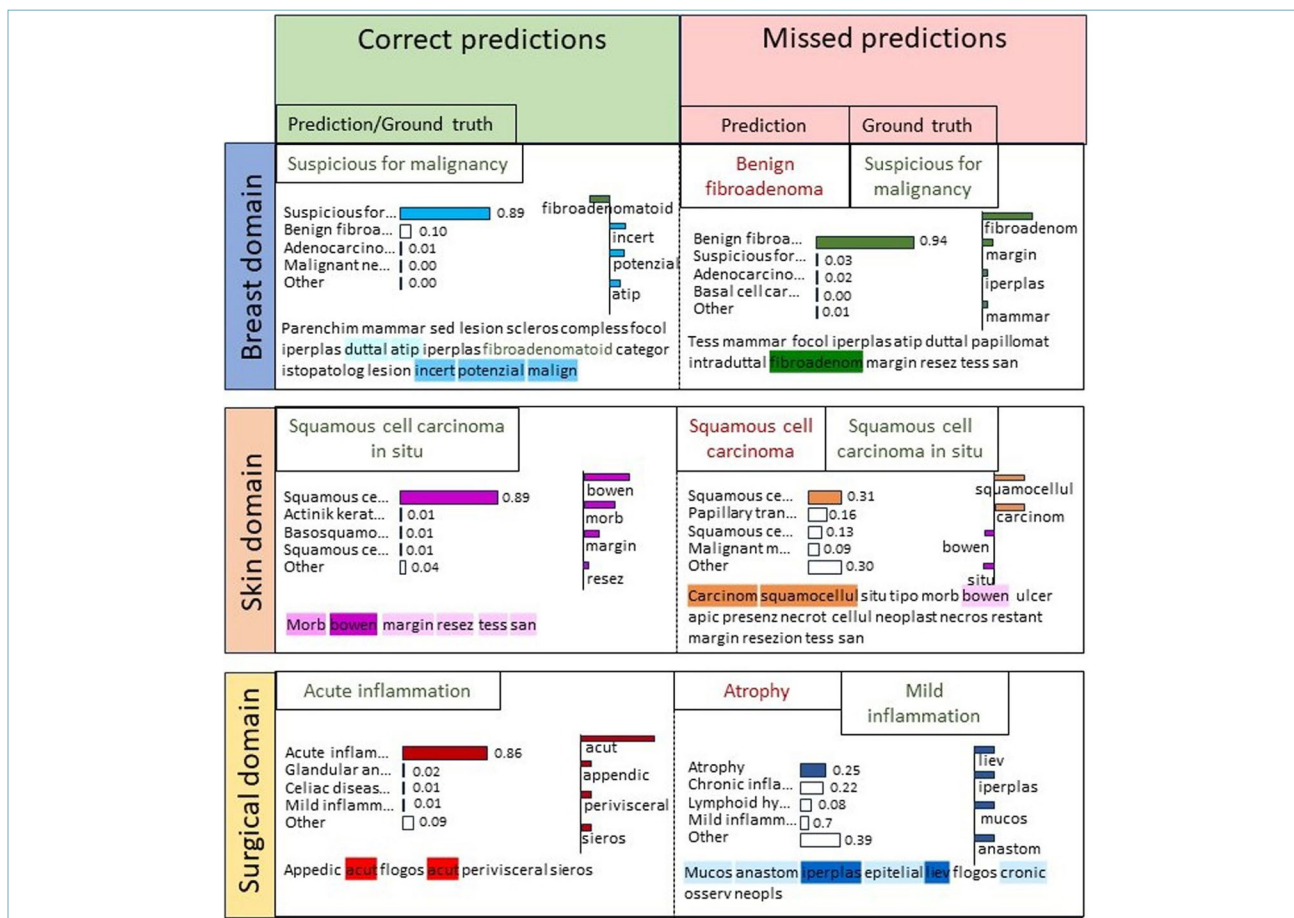
**Table I.** Comparison of metrics achievement between LSTM and SVM models. SVM outperformed LSTM achieving better results. LSTM, long short-term memory; SVM, support vector machine; SD, standard deviation; 95%CI, 95% confidence interval.

	LSTM	SVM	p-value
Accuracy	0.83	0.84	
Precision mean (SD) [95%CI]	0.85 (0.21) [0.19-1]	0.87 (0.22) [0.20-1]	0.64
Recall mean (SD) [95%CI]	0.83 (0.26) [0.15 -1]	0.83 (0.21) [0.21-1]	1
F1-Score mean (SD) [95%CI]	0.82 (0.26) [0.15-1]	0.82 (0.23) [0.21-1]	1

reasons for the missed ones as demonstrated in the examples reported in Figure 3. In particular, within different semantic domains (e.g. breast, skin and general pathology) the NLP algorithms demonstrated fluctuant accuracy in predicting the correct SNOMED-CT based on the availability of sufficient keywords for the correct assignment. As a result, cases of atypical fibroadenomatoid proliferations of the breast were correctly (Atypia, suspicious for malignancy) or incorrectly (Benign fibroadenoma) classified due to the presence/absence of concurrent keywords such as “atypical” or “unknown malignant potential”. The same was noted for in situ squamous lesions of the skin, where the term “Bowen” complicated the assignment task, and in inflammatory conditions where terms like “acute” or “chronic” were key for the correct SNOMED-CT assignment.

**Discussion**

One of the major challenges that our discipline is facing within the digital pathology revolution is linked to the archiving phase, and specifically for reports,



**Figure 3.** Examples of correct (left column) and missed (right column) SNOMED-CT prediction in different domains (breast, skin and surgical/general pathology are reported here). Each box reports the most probable SNOMED-CT code attributable to the case by the NLP model, the predicted code and the ground truth, the most relevant keywords used by the model to predict the code and examples of the narrative text with most relevant keywords highlighted in progressively dark colors.

whose standardized and systematic organization would allow efficient retrieval and storage of patient information<sup>17</sup>. The progressive prospective adoption of SSR, eventually associated with NLP able of extracting granular information from structured and narrative reports, will probably solve this issue when adequately integrated within our LIS and widely employed in our departments. However, while SSR are largely available and applicable in the oncological setting<sup>18,19</sup>, their application in other pathology fields (e.g. inflammatory conditions) is still delayed, due to the higher variability and complexity of these reports and to the lack of standardization<sup>18,20,21</sup>. Moreover, the extraction of specific features from structured reports through NLP can be computationally demanding. To overcome this challenge, restricting the research to smaller disease groups through appropriate coding (e.g. SNOMED-CT) can prospectively facilitate the

information extraction process. Finally, the complexity that our reports are reaching through the integration of molecular and genomic big data incorporation represent a further source of information overload, requiring a strong standardization effort (e.g. with health informatics standards as CDA or FHIR) ensuring that the data are computer identifiable, retrievable and processable. In this direction, the Italian Society of Pathology (SIAPEC) is working on the creation of a standard dataset for reporting, moving from free-text narrative reports (level 1 according to the Ontario scale)<sup>22</sup> to synoptic reports (level 3) to fully structured reporting, which should include discrete data embedded in LIS and structured messaging/data exchange standards (level 6), culminating in the creation of common and shared data lake<sup>23</sup>. Meanwhile, our archives are already hosting a large quantity of narrative reports whose retrieval is often inaccurate or com-

pletely lacking due to their intrinsic “analog” nature. In this setting, the employment of SNOMED-CT can represent a solution for labeling cases to help in their retrospective retrieval and organization<sup>24,25</sup>. Although limited by the partial adoption of SNOMED-CT among departments and its need for progressive updates (e.g. for molecular and omic terms)<sup>26,27</sup>, leveraging on these codes can potentially help us in re-organizing the retrospective pathology archives<sup>8,28,29</sup>. In this direction, the application of rule-based approaches have already been proposed to overcome the intrinsic limits (time-consuming process, prone to human error and interobserver variability) of manual SNOMED-CT coding<sup>30-32</sup>. The present study demonstrates that the application of SVM and LSTM AI tools can be a valid alternative to manual rule engineering, showing proficiency in capturing complex patterns that simpler techniques may overlook, improving automation in the setting of NLP, and providing a temporary fix while SSR are more widely adopted in our departments. Moreover, interpretability analysis highlighted the impact of specific words of the pathology report on the final decision. In particular, this visual analysis demonstrated the need for additional keywords that may contribute to the final assignment of a SNOMED-CT code together with the “defining” words of particular entities. In this setting, emblematic is the example of breast pathology domain, where the coexistence of uncertainty keywords (e.g. “atypical” or “unknown malignant potential”), along with specific terms indicating the disease (e.g. “fibroadenomatous”), were essential to avoid misassignment of the code (benign, fibroadenoma), allowing the correct identification of the case (Atypia, suspicious for malignancy). Finally, the presence of narrative/descriptive reports can be a source of variability in the final code assignment, as documented by the example reported in acute vs chronic inflammation and atrophy. The proposed NLP-based methods can thus represent a valid supporting tool for the retrospective labeling of narrative reports, waiting for a more comprehensive implementation of tandem SSR/NLP approach in our LIS. However, the further expansion of alternative language/vocabularies<sup>33</sup>, the fine-tuning of SVM to account variations across languages<sup>34</sup>, the extension to more “rare” codes that goes beyond the 50 tested here can help to overcome the limits of generalizability with rarer diagnosis, improving this tool for the actual employment by researchers and pathologists<sup>35</sup>. Challenges faced by SNOMED CT in its widespread use, such as redundant codes, manual errors, and difficulties in creating user-friendly labeling modules, also impact the dataset’s reliability.<sup>36</sup> The future implementation of automatic coding through LIS-integrated SSR promises a more robust

and reliable future dataset on which these algorithms can be fine-tuned<sup>23</sup>.

## Conclusions

The application of AI tools allows the automatic SNOMED-CT labeling of the pathology archives, providing a retrospective fix to the large lack of organization of narrative reports. This rescue solution can pave the way a the combined NLP and SSR prospective approach through adequate LIS implementation for improving patients’ care.

## ACKNOWLEDGEMENTS

We gratefully acknowledge Giuseppe Barletta for his contributions to this article, including the creation of the SVM model and valuable IT support.

## CONFLICTS OF INTEREST

Authors do not have any conflict of interest to disclose.

## FUNDING

The present work has been funded by the Italian Ministry of the University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337).

## AUTHORS’ CONTRIBUTIONS

GC defined the study design, performed the retrospective research and data extraction, as well as the elaboration of the computational pipeline for the study; APDT, MS, AG, AE, VDM, EM, MC, SM and EB provided counseling as experts in the nomenclature and SNOMED field; FP and VL performed the supervision of the work revising critically the manuscript before the approval by all the authors; FP provided the funding acquisition and administrative support. All authors were involved in writing the paper and had final approval of the submitted and published versions.

## ETHICAL CONSIDERATION

The research was conducted ethically, with all study procedures being performed in accordance with the requirements of the World Medical Association’s Declaration of Helsinki.

## References

- 1 Caputo A, L’Imperio V, Merolla F, et al. The slow-paced digital evolution of pathology: lights and shadows from a multifaceted board. *Pathologica* 2023;115:127-136. <https://doi.org/10.32074/1591-951X-868>
- 2 Cazzaniga G, Rossi M, Eccher A, et al. Time for a full digital approach in nephropathology: a systematic review of current artificial intelligence applications and future directions. *J Nephrol*

- Published online September 28, 2023. <https://doi.org/10.1007/s40620-023-01775-w>
- 3 Cazzaniga G, Bolognesi MM, Stefania MD, et al. Congo Red Staining in Digital Pathology: The Streamlined Pipeline for Amyloid Detection Through Congo Red Fluorescence Digital Analysis. *Lab Invest* 2023;103:100243. <https://doi.org/10.1016/j.labinv.2023.100243>
  - 4 L'Imperio V, Wulczyn E, Plass M, et al. Pathologist Validation of a Machine Learning-Derived Feature for Colon Cancer Risk Stratification. *JAMA Netw Open* 2023;6:e2254891. <https://doi.org/10.1001/jamanetworkopen.2022.54891>
  - 5 Marletta S, L'Imperio V, Eccher A, et al. Artificial intelligence-based tools applied to pathological diagnosis of microbiological diseases. *Pathol Res Pract* 2023;243:154362. <https://doi.org/10.1016/j.prp.2023.154362>
  - 6 L'Imperio V, Gibilisco F, Fraggetta F. What is Essential is (No More) Invisible to the Eyes: The Introduction of BlocDoc in the Digital Pathology Workflow. *J Pathol Inform* 2021;12:32. [https://doi.org/10.4103/jpi.jpi\\_35\\_21](https://doi.org/10.4103/jpi.jpi_35_21)
  - 7 Munari E, Scarpa A, Cima L, et al. Cutting-edge technology and automation in the pathology laboratory. *Virchows Arch* Published online November 6, 2023. <https://doi.org/10.1007/s00428-023-03637-z>
  - 8 Eccher A, Dei Tos AP, Scarpa A, et al. Cost analysis of archives in the pathology laboratories: from safety to management. *J Clin Pathol*. Published online August 2, 2023. <https://doi.org/10.1136/jcp-2023-209035>
  - 9 Eccher A, Scarpa A, Dei Tos AP. Impact of a centralized archive for pathology laboratories on the health system. *Pathol Res Pract* 2023;245:154488. <https://doi.org/10.1016/j.prp.2023.154488>
  - 10 Fraggetta F, L'Imperio V, Ameisen D, et al. Best Practice Recommendations for the Implementation of a Digital Pathology Workflow in the Anatomic Pathology Laboratory by the European Society of Digital and Integrative Pathology (ESDIP). *Diagnostics (Basel)* 2021;11:2167. <https://doi.org/10.3390/diagnostics11112167>
  - 11 Fraggetta F, Caputo A, Guglielmino R, et al. A Survival Guide for the Rapid Transition to a Fully Digital Workflow: The "Caltagirone Example." *Diagnostics (Basel)* 2021;11:1916. <https://doi.org/10.3390/diagnostics11101916>
  - 12 L'Imperio V, Casati G, Cazzaniga G, et al. Improvements in digital pathology equipment for renal biopsies: updating the standard model. *J Nephrol*. Published online February 14, 2023. <https://doi.org/10.1007/s40620-023-01568-1>
  - 13 L'Imperio V, Brambilla V, Cazzaniga G, Ferrario F, Nebuloni M, Pagni F. Digital pathology for the routine diagnosis of renal diseases: a standard model. *J Nephrol* 2021;34:681-688. <https://doi.org/10.1007/s40620-020-00805-1>
  - 14 Pallua JD, Brunner A, Zelger B, Schirmer M, Haybaeck J. The future of pathology is digital. *Pathol Res Pract* 2020;216:153040. <https://doi.org/10.1016/j.prp.2020.153040>
  - 15 Snoek JAA, Nagtegaal ID, Siesling S, van den Broek E, van Slooten HJ, Hugen N. The impact of standardized structured reporting of pathology reports for breast cancer care. *Breast* 2022;66:178-182. <https://doi.org/10.1016/j.breast.2022.10.011>
  - 16 Soysal E, Warner JL, Wang J, et al. Developing Customizable Cancer Information Extraction Modules for Pathology Reports Using CLAMP. *Stud Health Technol Inform* 2019;264:1041-1045. <https://doi.org/10.3233/SHT1190383>
  - 17 Gutman DA, Khalilia M, Lee S, et al. The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research. *Cancer Res* 2017;77:e75-e78. <https://doi.org/10.1158/0008-5472.CAN-17-0629>
  - 18 Ellis DW, Srigley J. Does standardised structured reporting contribute to quality in diagnostic pathology? The importance of evidence-based datasets. *Virchows Arch* 2016;468:51-59. <https://doi.org/10.1007/s00428-015-1834-4>
  - 19 Hewer E. The Oncologist's Guide to Synoptic Reporting: A Primer. *Oncology* 2020;98:396-402. <https://doi.org/10.1159/000500884>
  - 20 Leh S, Dendooven A. Systematic reporting of medical kidney biopsies. *Clin Kidney J* 2021;15:21-30. <https://doi.org/10.1093/ckj/sfab140>
  - 21 Langford CR, Goldinger MH, Treanor D, et al. Improved pathology reporting in NAFLD/NASH for clinical trials. *J Clin Pathol* 2022;75:73-75.
  - 22 Srigley JR, McGowan T, Maclean A, et al. Standardized synoptic cancer pathology reporting: a population-based approach. *J Surg Oncol* 2009;99:517-524. <https://doi.org/10.1002/jso.21282>
  - 23 Renshaw AA, Mena-Allauca M, Gould EW, et al. Synoptic Reporting: Evidence-Based Review and Future Directions. *CO Clin Cancer Inform* 2018;2. <https://doi.org/10.1200/CCI.17.00088>
  - 24 Paskal W, Paskal AM, Dębski T, Gryziak M, Jaworowski J. Aspects of Modern Biobank Activity - Comprehensive Review. *Pathol Oncol Res* 2018;24:771-785. <https://doi.org/10.1007/s12253-018-0418-4>
  - 25 Ali M, Evans H, Whitney P, Minhas F, Snead DRJ. Using Systemised Nomenclature of Medicine (SNOMED) codes to select digital pathology whole slide images for long-term archiving. *J Clin Pathol* 2023;76:349-352. <https://doi.org/10.1136/jcp-2022-208483>
  - 26 Ceusters W. SNOMED CT revisions and coded data repositories: when to upgrade? *AMIA Annu Symp Proc* 2011;2011:197-206.
  - 27 Jiang X. Intelligent Classification Method of Archive Data Based on Multigranular Semantics. *Comput Intell Neurosci* 2022;2022:7559523. <https://doi.org/10.1155/2022/7559523>
  - 28 Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. *J Am Med Inform Assoc* 2006;13:536-546. <https://doi.org/10.1197/jamia.M2093>
  - 29 Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008;8 Suppl 1(Suppl 1):S2. <https://doi.org/10.1186/1472-6947-8-S1-S2>
  - 30 Carter KJ, Rinehart S, Kessler E, et al. Quality assurance in anatomic pathology: automated SNOMED coding. *J Am Med Inform Assoc* 1996;3:270-272. <https://doi.org/10.1136/jamia.1996.96413134>
  - 31 García-Rojo M, Daniel C, Laurinavicius A. SNOMED CT in pathology. *Stud Health Technol Inform* 2012;179:123-140.
  - 32 Skeppstedt M, Kvist M, Dalianis H. Rule-based Entity Recognition and Coverage of SNOMED CT in Swedish Clinical Text. In: *LREC*; 2012:1250-1257.
  - 33 Millar J. The Need for a Global Language - SNOMED CT Introduction. *Stud Health Technol Inform*. 2016;225:683-685.
  - 34 Schulz S, Hammer L, Nik DH, Kreuzthaler M. Localising the Clinical Terminology SNOMED CT by Semi-automated Creation of a German Interface Vocabulary. *MULTILINGUALBIO*. Published online 2020. Accessed June 4, 2023. <https://www.semanticscholar.org/paper/92fa8856652d1a7cf2d4781da45e4829fd0a2fd6>
  - 35 Amin M, Dhir R. Data Representation, Coding, and Communication Standards. *Surg Pathol Clin* 2015;8:109-121. Amin M, Dhir R. Data Representation, Coding, and Communication Standards. *Surg Pathol Clin* 2015;8:109-121.
  - 36 Deeken-Draisey A, Ritchie A, Yang GY, et al. Current Procedural Terminology Coding for Surgical Pathology: A Review and One Academic Center's Experience With Pathologist-Verified Coding. *Arch Pathol Lab Med* 2018;142:1524-1532. <https://doi.org/10.5858/arpa.2017-0190-RA>