

Response to Kempf et al on Methodological and Practical Aspects of a Distant Metastasis Detection Model

We have carefully read the letter “The More, the Better? Modalities of Metastatic Status Extraction on Free Medical Reports Based on Natural Language Processing” sent to *JCO Clinical Cancer Informatics* by Kempf et al,¹ concerning our publication “Automatic Detection of Distant Metastasis Mentions in Radiology Reports in Spanish.”² We appreciate the interest in our paper and in pointing out aspects of great relevance today, especially the sustainability of machine learning models because of their high carbon footprint.

As we highlight in our publication, context is essential when it comes to determining the metastatic status of a tumor since this status depends on the primary tumor site and the anatomic site of the secondary lesion. In this sense, a rule model can provide good results, as demonstrated by Kempf et al³ and Guével et al,⁴ who by using a rule-based algorithm obtained a recall, precision, and F1 score of 0.80, 0.75, and 0.79 on 227 colon and breast cancer reports. However, rule-based methods are limited, especially in imaging reports in Spanish, which show fundamental characteristics that influence the assignment of the M1 or M0 label to an entity. As we noted in our publication, the Oncology Institute Fundación Arturo López Pérez—Radiology Report corpus presented a total of 7,479 annotated entities, of which 5,640 were M0 and 1,839 were M1 (approximate ratio of 3:1). The difference between these two entities is that the M0 are mentions of metastasis with modifiers (negation and uncertainty), which are located at very variable distances from the mention of interest, such as *se descarta* (is discarded), *sin evidencia de* (without evidence of), *sin* (without), or *de significado incierto* (of uncertain meaning). By contrast, M1 mentions lack modifiers and are affirmative.

Beyond additional observations that make comparisons between performances concerning our model unfair (different languages, models trained on different types of cancer, comparison of a model trained on two cancers—Kempf’s model—versus our multiple models trained on one cancer, etc), it is valid to highlight that the comparison with a base model must be reported.

Our baseline used a string match algorithm between a lexicon with the words or phrases associated with metastatic lesions and the tokenized phrases from the reports. The NegEx-MES algorithm was used to determine whether the mention was negated or not in the context of the phrase that contained it. This model allowed the reports to be classified as M0 or M1 with a recall, precision, and F1 score of 0.82, 0.85, and 0.81 (Table 1).

In comparison, our ML-based model achieved a recall, precision, and F1 score of 0.92, 0.88, and 0.91, respectively. This improvement of 0.1 in the F1 score demonstrates a significant advantage in the overall performance of the ML model, justifying, at least in part, its use despite the higher computational cost. This performance gain is critical in clinical settings where accurate and reliable identification of metastatic status can significantly affect patient outcomes.

Regarding the annotation of the corpus, as we also described in our publication, the annotation process did not involve only two general nursing technicians. The two annotators, in addition to being experts in oncological registries, had previous experience in annotation and coding of medical entities. Agile annotation cycles⁵ were used, with weekly meetings to analyze cases of disagreement jointly with a radiologist and a nuclear medicine physician. Furthermore, once the annotation campaign was finished, the manager analyzed each annotation to standardize the corpus. The consistency of the annotation was also evaluated in 20% of the corpus with high agreement between annotators. Both the final version of the corpus⁶ and the annotation guidelines⁷ are available for analysis.


Although the personnel involved were not exclusively medical experts with annotation training, the procedure was methodical and rigorous and allowed the development of a tool to accelerate the identification of the metastatic status by clinicians, who have the final decision in the classification of patients. Recently it has been suggested that on the way to building gold standards, the manual annotation process can be accelerated by using masked language models and causal language models as entity preannotation methods, especially the latter, since they have lower performance in information extraction, while the former have better performance in terms of higher F1 and lower CO₂ emissions.⁸ Of course, the new opportunities that are opening up the large language models must be weighed against the potential environmental impact they could bring.

TABLE 1. Performance Metrics for the Document-Level Classification Model Based on String Match and NegEx-MES


Entity	Recall	Precision	F1 Score	Support
M0	0.62	0.95	0.75	27
M1	0.97	0.77	0.86	40
Mean average	0.82	0.85	0.81	67

Corresponding author: Pablo Báez, PhD; e-mail: pf.baez@udd.cl.

We are committed to continuously improving our model and methodology, and we welcome further collaboration and discussion with the research community to address these challenges and advance the field.

Ricardo Ahumada, MSc 

Center of Medical Informatics and Telemedicine, Faculty of Medicine, Universidad de Chile, Santiago, Chile

Jocelyn Dunstan, PhD 

Department of Computer Science & the Institute for Mathematical Computing, Pontificia Universidad Católica de Chile, Santiago, Chile

Inti Paredes, MD

Departamento de Investigación del Cáncer, Unidad de Informática Médica y Data Science, Instituto Oncológico Fundación Arturo López Pérez, Santiago, Chile

Pablo Báez, PhD 

Center of Medical Informatics and Telemedicine, Faculty of Medicine, Universidad de Chile, Santiago, Chile

Tecnología Médica, Facultad de Medicina, Universidad del Desarrollo, Santiago, Chile

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Inti Paredes

Employment: Arturo Lopez Perez Foundation

Leadership: Arturo Lopez Perez Foundation

No other potential conflicts of interest were reported.

REFERENCES

- Kempf E, Priou S, Cohen A, et al: The more, the better? Modalities of metastatic status extraction on free medical reports based on natural language processing. *JCO Clin Cancer Inform* [10.1200/CCI.24.00026](https://doi.org/10.1200/CCI.24.00026)
- Ahumada R, Dunstan J, Rojas M, et al: Automatic detection of distant metastasis mentions in radiology reports in Spanish. *JCO Clin Cancer Inform* [10.1200/CCI.23.00130](https://doi.org/10.1200/CCI.23.00130)
- Kempf E, Priou S, Lamé G, et al: Impact of two waves of Sars-Cov2 outbreak on the number, clinical presentation, care trajectories and survival of patients newly referred for a colorectal cancer: A French multicentric cohort study from a large group of university hospitals. *Int J Cancer* 150:1609-1618, 2022
- Guével E, Priou S, Lamé G, et al: Impact of the COVID-19 pandemic on clinical presentation, treatments, and outcomes of new breast cancer patients: A retrospective multicenter cohort study. *Cancer Med* 12:20918-20929, 2023
- Fort K: Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects. Hoboken, NJ, John Wiley & Sons, 2016
- Ahumada R, Báez P, Paredes I, et al: FALP Radiology reports: Annotated corpus for distant metastasis detection (1.0) [data set]. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7623396>
- Ahumada R, Báez P, Caamaño BG, et al: Annotation guidelines for FALP radiology reports annotated corpus for distant metastasis detection (1.1) [data set]. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7623509>
- Naguib M, Tannier X, Névéol A. Few shot clinical entity recognition in three languages: Masked language models outperform LLM prompting. arXiv preprint arXiv:2402.12801, 2024

DOI: <https://doi.org/10.1200/CCI-24-00154>; Published at ascopubs.org/journal/cci on August 26, 2024.