Letter to the Editor

Baseline Fidelity and Evaluation Independence in Symptom Terminology Normalization

Qi He, DM

Guangxi Hospital Division of The First Affiliated Hospital, Sun Yat-sen University, Nanning, Guangxi Zhuang Autonomous Region, China

Corresponding Author:

Qi He, DM
Guangxi Hospital Division of The First Affiliated Hospital
Sun Yat-sen University
No. 3 Foziling Road
Nanning, Guangxi Zhuang Autonomous Region 530022
China
Phone: 86 15521501876
Email: hqqqq2024@163.com

Related Articles:
Comment on: https://medinform.jmir.org/2026/1/e115495
Comment on: https://medinform.jmir.org/2026/1/e85825

JMIR Med Inform 2026;14:e113829; doi: 10.2196/113829

Keywords: symptom terminology; traditional Chinese medicine; natural language processing; named entity recognition; named entity normalization; pretrained language model

The split-then-concatenate normalization framework (STCNF) proposed by Yao et al [1] addresses the complexity of expressions in symptom terminology. Two evaluation issues warrant clarification before attributing the reported gains to the framework.

First, the paper characterizes a pretrained language model for *International Classification of Diseases* coding (PLM-*ICD*) as a single-output classifier and attributes its 0/369 exact-match accuracy for multi-implication expressions to this limitation [1]. However, the cited PLM-*ICD* formulation predicted multiple labels using label-specific sigmoid probabilities, binary cross-entropy, and threshold-based selection [2]. Could the authors specify whether their implementation retained this formulation or introduced a single-label adaptation? Reporting the output layer, loss function, decision rule, and threshold-selection procedure would resolve this discrepancy. Under exact-set accuracy, a single-output system necessarily fails whenever the reference contains multiple terms. Its zero accuracy therefore cannot establish an inherent limitation of the original PLM-*ICD* architecture. A multilabel implementation trained and evaluated on the same task would be a more informative comparator. This concern does not, by itself, explain STCNF's reported advantage over a multitask candidate generator (MTCG), which supports multiple outputs.

Second, the 400/100 electronic medical record split coexisted with the construction of infrequent-entity knowledge bases using all 500 records [1]. A fixed external terminology is distinct from a corpus-derived resource informed by the held-out records. The latter warrants clarification of the intended evaluation setting and possible information transfer across partitions [3]. The authors should specify whether these resources affected the outcomes in Table 6 or only auxiliary semantic attributes and document the provenance of the 3350 alignment pairs, including augmented examples. Were all corpus-derived training pairs generated exclusively from training records? The discussion also justifies selecting the entity-recognition component using the reported recognition results [1]. Was this choice prespecified or based on independent development data? The published description does not establish whether supervised test-label leakage occurred, nor quantify any resulting optimism.

A focused reanalysis could preserve externally specified terminology while restricting corpus-derived resources and training examples to the training partition, using separate development data for selection, and evaluating the frozen pipeline on untouched records. Reporting results separately for previously observed and unseen surface forms would further clarify generalization. Alongside a task-compatible PLM-*ICD* comparator, component ablations could distinguish the contributions of decomposition, semantic matching, and rule-based reconstruction. These clarifications would

strengthen interpretation of STC-NF without discounting its potentially useful approach.

Generative AI Use

ChatGPT (OpenAI) assisted with language editing.

Funding

No external funding was received for this work.

Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study

Authors' Contributions

QH contributed to conceptualization, methodology, investigation, writing (original draft), and writing (review and editing).

Conflicts of Interest

None declared.

References

1. Yao J, Gou X, Lai W, et al. Symptom terminology normalization in traditional Chinese medicine: development and evaluation of a 2-stage deep learning framework based on fine-grained semantic classification. JMIR Med Inform. Sep 25, 2026;14:e85825. [doi: 10.2196/85825] [Medline: 42789749]
2. Huang CW, Tsai SC, Chen YN. PLM-ICD: automatic ICD coding with pretrained language models. In: Proceedings of the 4th Clinical Natural Language Processing Workshop. Association for Computational Linguistics; 2022:10-20. [doi: 10.18653/v1/2022.clinicalnlp-1.2]
3. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns (NY). Sep 8, 2023;4(9):100804. [doi: 10.1016/j.patter.2023.100804] [Medline: 37720327]

Abbreviations

ICD: International Classification of Diseases
MTCG: multitask candidate generator
PLM-ICD: pretrained language model for International Classification of Diseases coding
STC-NF: split-then-concatenate normalization framework

Edited by Andrew Coristine; This is a non–peer-reviewed article; submitted 07.Oct.2026; initial comments to author 06.Nov.2026; final revised version received 07.Nov.2026; accepted 12.Nov.2026; published 26.Nov.2026

Please cite as:
He Q
Baseline Fidelity and Evaluation Independence in Symptom Terminology Normalization
JMIR Med Inform 2026;14:e113829
URL: https://medinform.jmir.org/2026/1/e113829
doi: 10.2196/113829

© Qi He. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.