International Journal of Population Data Science

Journal Website: www.ijpds.org

Mapping Free-Text Emergency Department Diagnoses to ICD-10 Codes Using Large Language Model Embeddings

Thomas Hartka[1], Robin Alexander[2], Rajendra Raman[3], Eilidh Mulhern[4], and Colin McCowan[2]

[1]University of Virginia, Charlottesville, USA
[2]University of St. Andrews, St. Andrews, United Kingdom
[3]Emergency Department, Victoria Hospital, Kirkcaldy, United Kingdom
[4]University of Glasgow, Glasgow, United Kingdom

Objective

In many emergency departments (EDs) in Scotland, discharge diagnoses are documented as free-text rather than coded using the International Classification of Diseases, Tenth Revision (ICD-10). This study aimed to validate an automated approach for mapping free-text ED discharge diagnoses to ICD-10 codes using large language model (LLM)–based text embeddings.

Methods

Data were obtained from the Health Informatics Centre for the years 2018–2021. From this data, 90 free-text ED discharge diagnoses were randomly selected. Three physicians independently assigned the ICD-10 code they felt best matched each diagnosis, with disagreements adjudicated by a fourth reviewer to establish a reference standard. Free-text diagnoses and ICD-10 code descriptions were embedded using the Hugging Face all-MiniLM-L6-v2 sentence-transformer model. For each diagnosis, the closest ICD-10 match was identified using cosine similarity. The accuracy of each physician reviewer and of the embedding-based approach, relative to the adjudicated reference standard, was assessed at increasing levels of ICD-10 specificity.

Results

Physician agreement with adjudicated codes ranged from 93–96% at the first character, 91–92% at the first two characters, 88–90% at the first three characters, and 76–79% for the full ICD-10 code. The embedding-based approach achieved accuracies of 89%, 86%, 83%, and 68% at the corresponding levels of specificity.

Conclusion

Mapping free-text ED diagnoses to ICD-10 codes is challenging and subject to inter-physician variability. While the embedding-based approach did not match physician-level performance, it demonstrated reasonable accuracy and may support scalable classification of ED diagnoses in large administrative databases.

https://doi.org/10.23889/ijpds.v11i5.3730