

Accepted Manuscript

Deep Neural Models for ICD-10 Coding of Death Certificates and Autopsy Reports in Free-Text

Francisco Duarte, Bruno Martins, Cátia Sousa Pinto, Mário J. Silva

PII: S1532-0464(18)30030-3
DOI: <https://doi.org/10.1016/j.jbi.2018.02.011>
Reference: YJBIN 2933

To appear in: *Journal of Biomedical Informatics*

Received Date: 18 September 2017
Accepted Date: 19 February 2018



Please cite this article as: Duarte, F., Martins, B., Pinto, C.S., Silva, M.J., Deep Neural Models for ICD-10 Coding of Death Certificates and Autopsy Reports in Free-Text, *Journal of Biomedical Informatics* (2018), doi: <https://doi.org/10.1016/j.jbi.2018.02.011>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Deep Neural Models for ICD-10 Coding of Death Certificates and Autopsy Reports in Free-Text

Francisco Duarte^a, Bruno Martins^a, Cátia Sousa Pinto^b, Mário J. Silva^a

^a*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

^b*Direção-Geral da Saúde, Portugal*

Abstract

We address the assignment of ICD-10 codes for causes of death by analyzing free-text descriptions in death certificates, together with the associated autopsy reports and clinical bulletins, from the Portuguese Ministry of Health. We leverage a deep neural network that combines word embeddings, recurrent units, and neural attention, for the generation of intermediate representations of the textual contents. The neural network also explores the hierarchical nature of the input data, by building representations from the sequences of words within individual fields, which are then combined according to the sequences of fields that compose the inputs. Moreover, we explore innovative mechanisms for initializing the weights of the final nodes of the network, leveraging co-occurrences between classes together with the hierarchical structure of ICD-10. Experimental results attest to the contribution of the different neural network components. Our best model achieves accuracy scores over 89%, 81%, and 76%, respectively for ICD-10 chapters, blocks, and full-codes. Through examples, we also show that our method can produce interpretable results, useful for public health surveillance.

Keywords: Automated ICD Coding, Clinical Text Mining, Deep Learning, Natural Language Processing, Artificial Intelligence in Medicine

Email addresses: francisco.ribeiro.duarte@tecnico.ulisboa.pt (Francisco Duarte), bruno.g.martins@tecnico.ulisboa.pt (Bruno Martins), catiasousapinto@dgs.min-saude.pt (Cátia Sousa Pinto), mario.gaspar.silva@tecnico.ulisboa.pt (Mário J. Silva)

Download English Version:

<https://daneshyari.com/en/article/6927490>

Download Persian Version:

<https://daneshyari.com/article/6927490>

[Daneshyari.com](https://daneshyari.com)