Transfer Learning for Early Prediction of Adverse Drug Reactions: Docetaxel and Alopecia in Breast Cancer as a Case Study

Stella DIMITSAKIa,b,1, Pantelis NATSIAVASb and Marie-Christine JAULENTa

a Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d’Informatique Médicale et d’Ingénierie des Connaissances pour la santé, LIMICS, F-75006 Paris, France
b Institute of Applied Biosciences, Centre for Research and Development Hellas, Thessaloniki, Greece

ORCiD ID: Stella Dimitsaki https://orcid.org/0000-0002-6356-9531
Pantelis Natsiavas https://orcid.org/0000-0002-4061-9815
Marie-Christine Jaulent https://orcid.org/0000-0003-4445-7494

Abstract. Transfer Learning (TL) is an approach which has not yet been widely investigated in healthcare, mostly applied in image data. This study outlines a TL pipeline leveraging Individual Case Safety reports (ICSRs) and Electronic Health Records (EHR), applied for the early detection Adverse Drug Reactions (ADR), evaluated using of alopecia and docetaxel on breast cancer patients as use case.

Keywords. Breast cancer, docetaxel, adverse drug reaction, alopecia, transfer learning

1. Introduction

Transfer Learning (TL) is a Machine Learning (ML) approach where a trained model for a specific task/dataset is retrained for another task/dataset. In healthcare, TL is currently mainly used for prediction models built upon image data [1]. Recently, TL is introduced also for Adverse Drug Reaction (ADR) prediction [2]. Our study aims to develop a TL predictive approach upon Individual Case Safety reports (ICSRs) and Electronic Health Records (EHR), applied for the early detection of ADRs signals, typically elaborated through disproportionality analysis applied in Pharmacovigilance (PV) pipelines. The presented approach is evaluated using alopecia, an ADR observed following chemotherapy with docetaxel in BC patients. An early detection of this ADR could help BC patients to confront anxiety and distress via proper psychological preparation and other potential mitigation measures (e.g., scalp cooling) [3].

1 Corresponding Author: Stella Dimitsaki, E-mail: sdimitsaki@certh.gr.
2. Methods

ICSR data from the FDA Adverse Event Reporting System (FAERS) will be used for the initial training step and the retrained step will be applied upon EHR data. Finally, the trained TF model will be evaluated in terms of predicting the adverse effect of alopecia on BC patients or the identification of the respective PV signal using the combination of ICSR and EHR data, compared to the use of tradition PV methods.

![Methodology rationale](image1.jpg)

**Figure 1.** Methodology rationale.

ICSR data are preprocessed, i.e. “cleaned” to facilitate further processing. Relevant information provided in ICSR data is taken into account, e.g. age and suspected drugs. Indicatively, Figure 2 depicts the number of concomitant suspected active substances in the selected ICSRs. Furthermore, ATC codes and the relevant classification of active substances is used to identify potential associations for drug classes.

![Top 26 concomitant/suspected active substances in the Docetaxel's ICSRs in FAERS](image2.jpg)

**Figure 2.** Top 26 concomitant/suspected active substances in the Docetaxel's ICSRs in FAERS

3. Conclusions

Based also on the outcome of other works combining both ICSRs and EHR data [4], we argue that such an “intelligent” approach could provide speeding the detection of ADRs.

References