Investigating ADR mechanisms with Explainable AI: a feasibility study with knowledge graph mining

Emmanuel Bresso[1,2], Pierre Monnin[1,2], Cédric Bousquet[2], François-Elie Calvier[2], Ndeye-Coumba Ndiaye[2], Nadine Petitpain[2], Malika Smaïl-Tabbone[1] and Adrien Coulet[1,2]
[1] Université de Lorraine, CNRS, Inria, LORIA, Nancy, France
[2] Please see [1] for full author information

Corresponding author: adrien.coulet@inria.fr

Reference paper: **Bresso *et al.* (2021) Investigating ADR mechanisms with Explainable AI: a feasibility study with knowledge graph mining. *BMC Medical Informatics and Decision Making.*** https://doi.org/10.1186/s12911-021-01518-6

Adverse drug reactions (ADRs) are statistically characterized within randomized trials or by post-marketing pharmacovigilance. However, the molecular mechanisms causing ADRs remain unknown in most cases. This is true even for common toxicities that are classically monitored during trials such as hepatic or skin toxicities. Interestingly, many elements of knowledge about drugs and drug ingredients are available beside clinical trials. In particular, open-access knowledge graphs describe their properties, interactions, and involvements in pathways. Expert classifications have also been manually established by experts and label drugs either as causative or not for several types of ADRs.

In our paper [1], we propose to mine biomedical knowledge graphs to identify biomolecular features that enable to automatically reproduce such expert classifications, distinguishing drugs causative or not for a given type of ADR. In an Explainable AI perspective, we explore simple classification techniques such as Decision Trees and Classification Rules because they provide human-readable models which explain the classification itself. We also evaluate the assumption that biomolecular features mined from knowledge graphs might provide elements of explanation for the molecular mechanisms behind ADRs. We tested our approach with two expert classifications identifying drugs causing or not hepatic or skin toxicities (respectively named *DILI* and *SCAR* for Drug Induced Liver Injuries and Severe Cutaneous Adverse Reactions). Features associated with these drugs were mined from PGxLOD [2], a biomedical knowledge graph that we previously created by interlinking public open data (including DisGeNET, PharmGKB, DrugBank, CTD). To this aim, we developed the `kgpm` algorithm [3] that enables scaling the extraction of paths of features up to a length of 4. Such paths are subsequently generalized into path patterns to cover larger sets of drugs. We trained two classifiers at distinguishing, on the basis of extracted features, drugs causing or not each of the two considered ADRs. We isolated features that are both discriminative in reproducing expert classifications and interpretable by experts (*e.g.*, *Gene Ontology* terms, drug targets, pathways) and asked 3 pharmacology experts to manually evaluate if they are potentially explanatory for ADRs.

Extracted features reproduce DILI and SCAR classifications with a fair fidelity (accuracy of 0.74 and 0.81, respectively). Experts fully agreed that 73% and 38% of the most discriminative features are possibly explanatory for DILI and SCAR, respectively; and partially agreed (2/3) for 90% and 77% of them. To illustrate, the path pattern $\xrightarrow{\text{interactsWith}}$ Enzyme $\xrightarrow{\text{cellularComponent}}$ Endoplasmic reticulum reached an agreement for DILI. Indeed, the endoplasmic reticulum is known, in particular in liver tissues, to host primarily cytochrome P450 enzymes, well known for being involved in drug metabolism.

From these results, it appears that the considered knowledge graph provides sufficiently diverse features to enable simple and explainable models to distinguish between drugs that are causative or not for ADRs. In addition to explaining expert classifications, most discriminative features appear to be good candidates for investigating further molecular mechanisms behind ADRs.

References

[1] Emmanuel Bresso et al. Investigating ADR mechanisms with explainable AI: a feasibility study with knowledge graph mining. *BMC Medical Informatics Decision Making*, 21(1):171, 2021.

[2] Pierre Monnin et al. PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison. *BMC Bioinformatics*, 20-S(4):139:1–139:16, 2019.

[3] Pierre Monnin et al. Tackling scalability issues in mining path patterns from knowledge graphs: a preliminary study. In *First International Conference "Algebras, graphs and ordered sets", ALGOS*, 2020.