

DeepARV-ChemBERTa: Ensemble Deep Learning to Predict Drug-Drug Interaction of Clinical Relevance with Antiretroviral Therapy



UNIVERSITY OF
LIVERPOOL

Thao.Pham@liverpool.ac.uk

Thao Pham¹, Mohamed Ghafoor¹, Sandra Grañana-Castillo¹, Catia Marzolini^{1,2}, Sara Gibbons¹, Saye Khoo¹, Justin Chiong¹, Saye Khoo¹, Dennis Wang^{3,4}, **Marco Siccardi^{1,5}**

¹University of Liverpool, Liverpool, UK, ²University Hospital Basel, Switzerland, ³Imperial College London, UK, ⁴Technology and Research (A*STAR), Republic of Singapore, ⁵esqLAB

Introduction

◆ **Drug-drug interaction (DDI)** may result in toxicity or treatment failure of antiretroviral therapy (ARV) or comedications [1]. Despite the high number of possible drug combinations in clinical use, only a limited number of clinical DDI studies are conducted before and after marketing approval.

◆ The cutting-edge approach is via **deep learning** [2], a sub-field of machine learning which is inspired by the human neural network. **Transformer-based model** is a type of neural network architecture that learns context and semantic information within sequential data, gaining significant popularity for its effectiveness in capturing complex patterns/relationships.

◆ Designed for applications in cheminformatics, **ChemBERTa** [3], pre-trained on 10 million compounds from PubChem, holds potential for further refinement for DDI prediction tasks.

◆ **Aims:** to build and assess the potential of our **deep learning approach, called DeepARV-ChemBERTa, to predict DDIs of clinical relevance** between ARVs and comedications.

Database

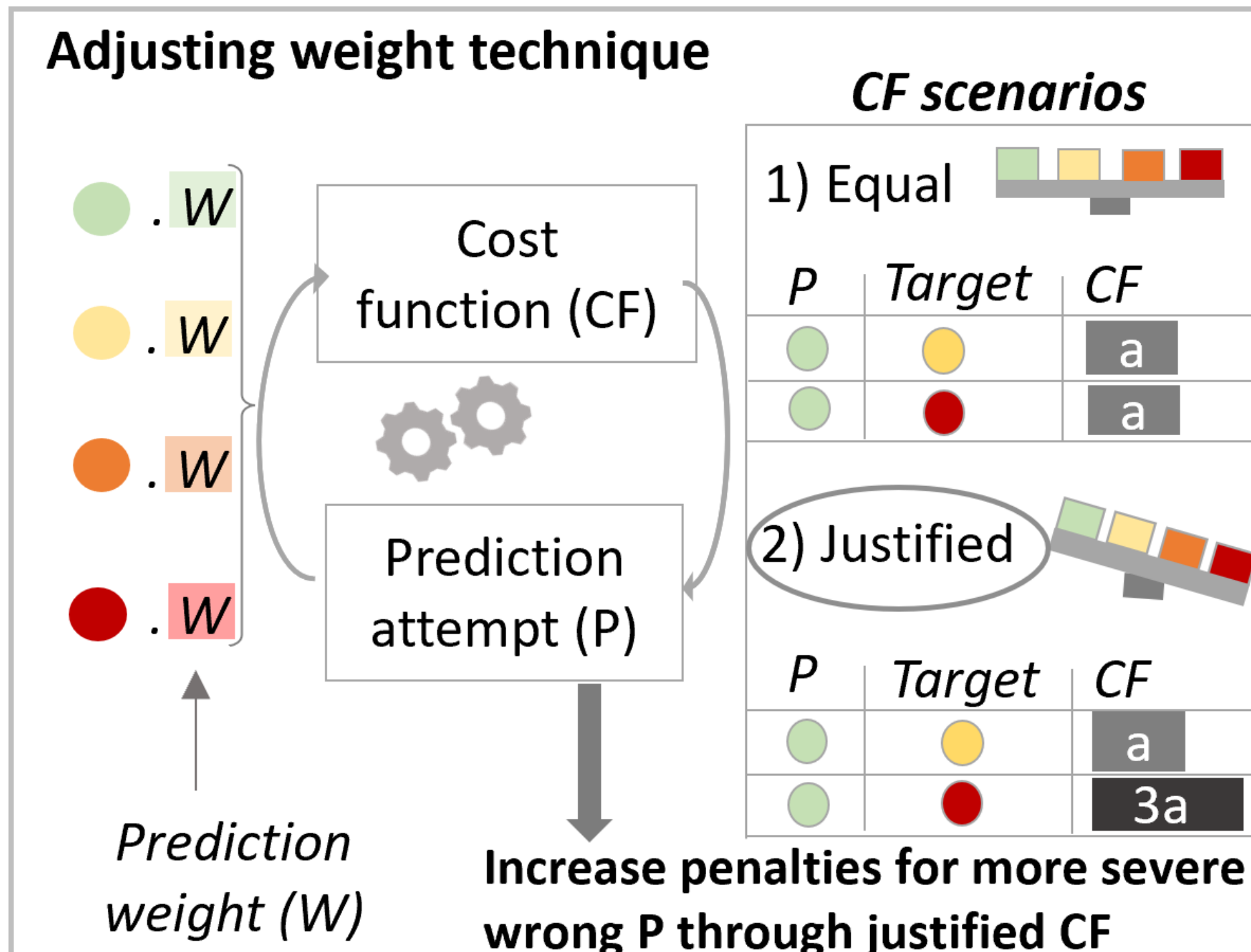
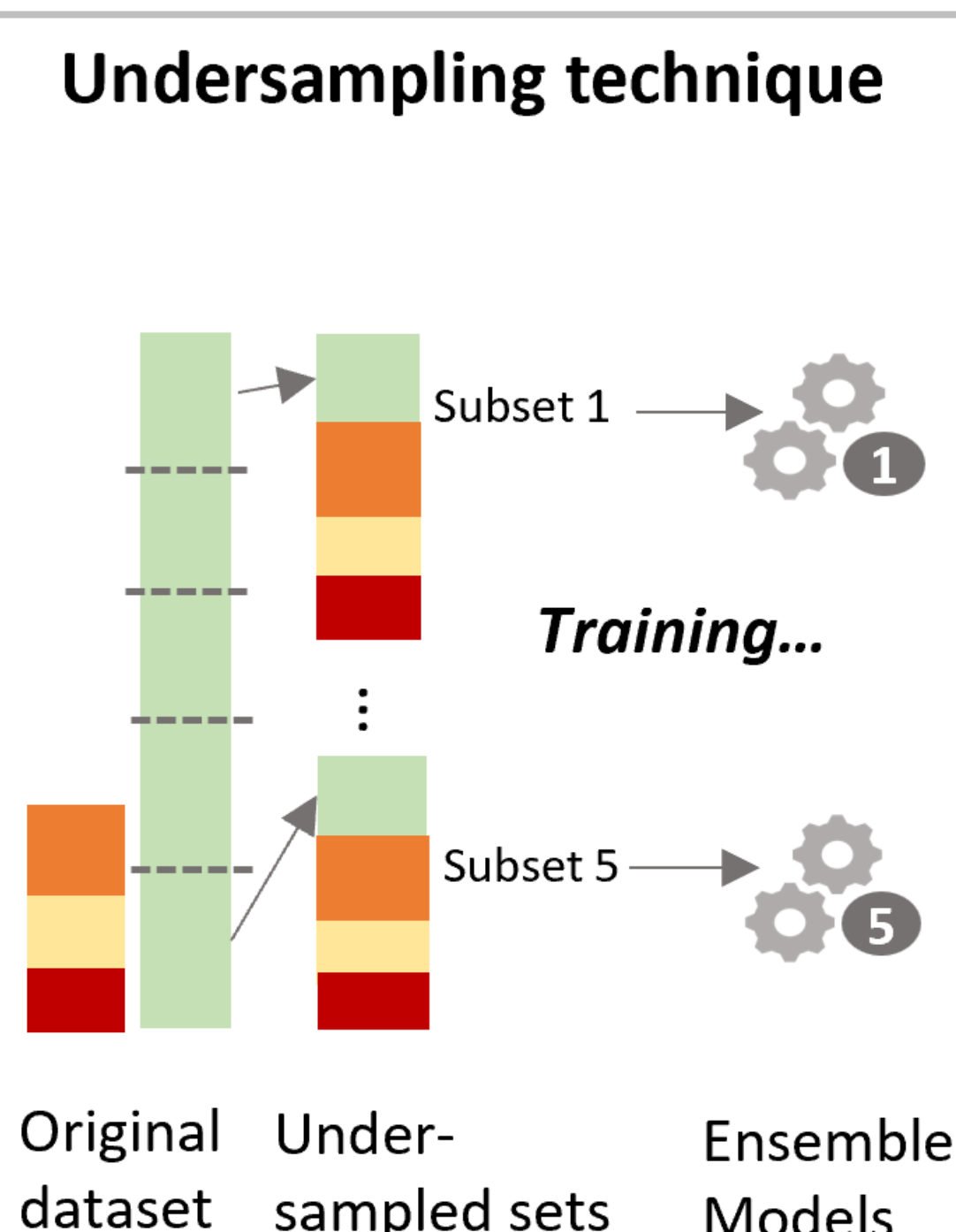
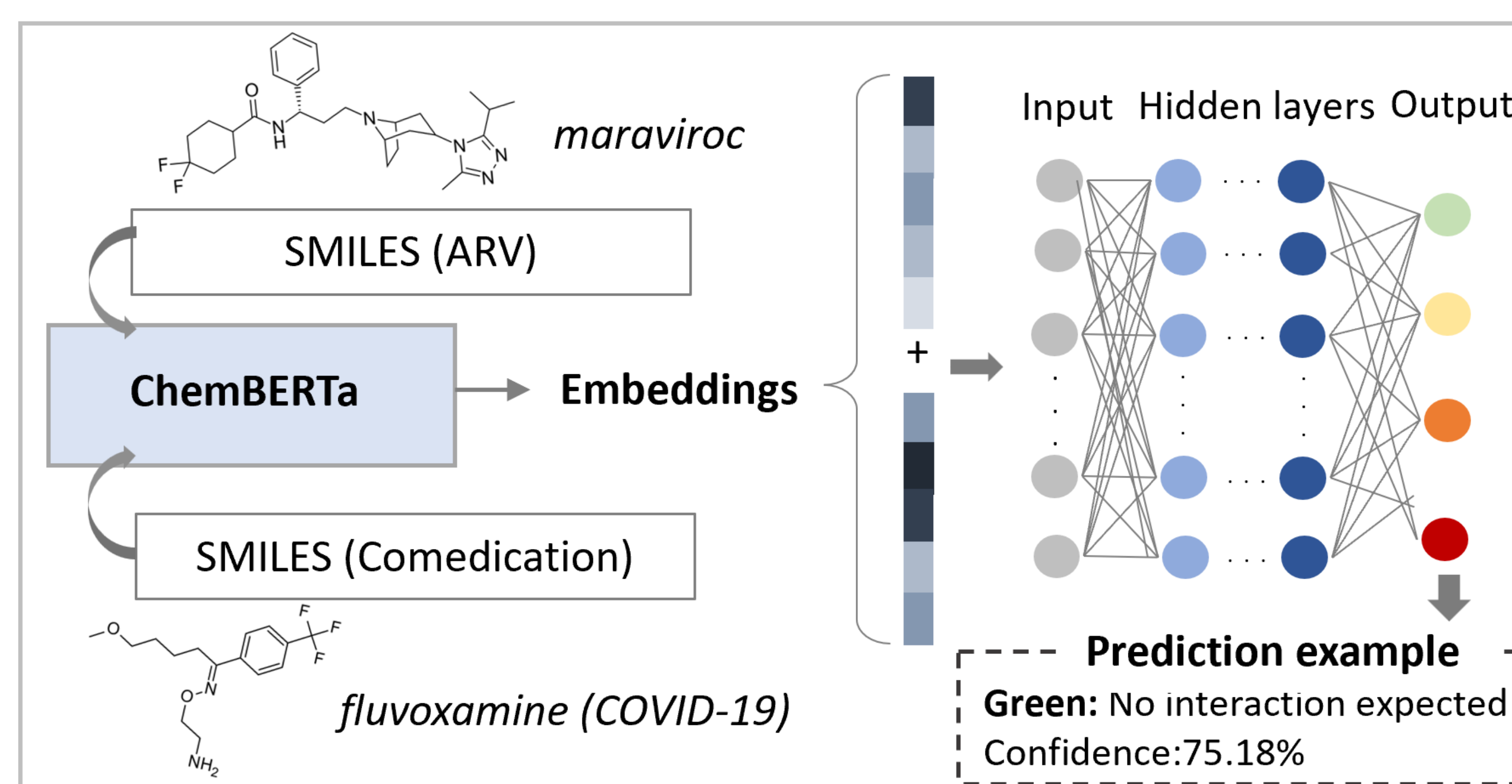
- ◆ Liverpool HIV Drug Interaction [4]
 - Primary ARVs: 42 drugs (single and combination)
 - Comedication: 687 drugs
 - Total: 28,026 drug pairs

Colour	DDI label
Green	No clinically significant interaction
Yellow	Potential interaction likely to be of weak interaction
Amber	Potential interaction that may require close monitoring
Red	Drugs should not be co-administered

Methods

A) Fine-tune the advanced transformer-based model, ChemBERTa-2 as a molecular feature extractor for the prediction of clinical DDI risks.

B) Employ downsampling and algorithmic adjusted weight techniques during training to address the imbalance challenge formed by the skewed distribution of DDI risks.



Results

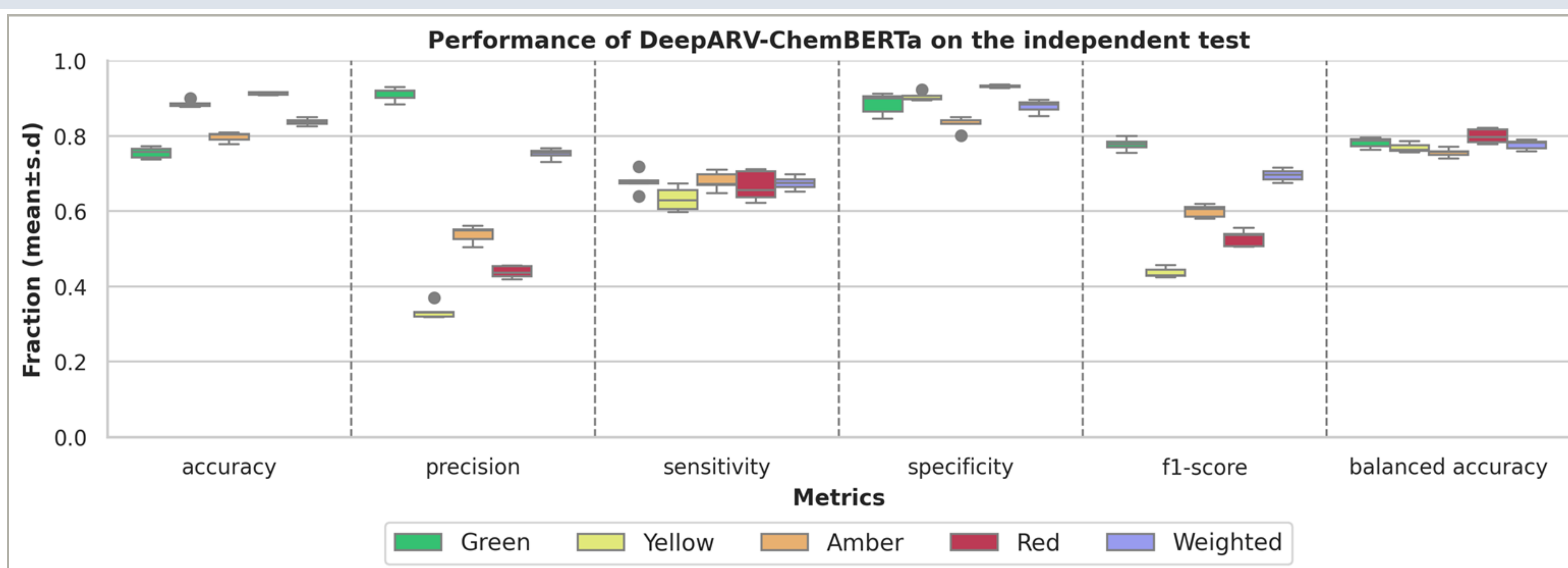


Fig 1. Performance of DeepARV-ChemBERTa on the independent test set. All trained models from 5-fold cross-validation were tested on the independent dataset with a total of 5,103 DDIs, which was composed of 3,221 Green, 358 Yellow, 1,146 Amber and 378 Red. The box plot showed the distribution of the performance scores across accuracy, precision, sensitivity, specificity, f1-score and balanced accuracy of each DDI category and the weighted macros.

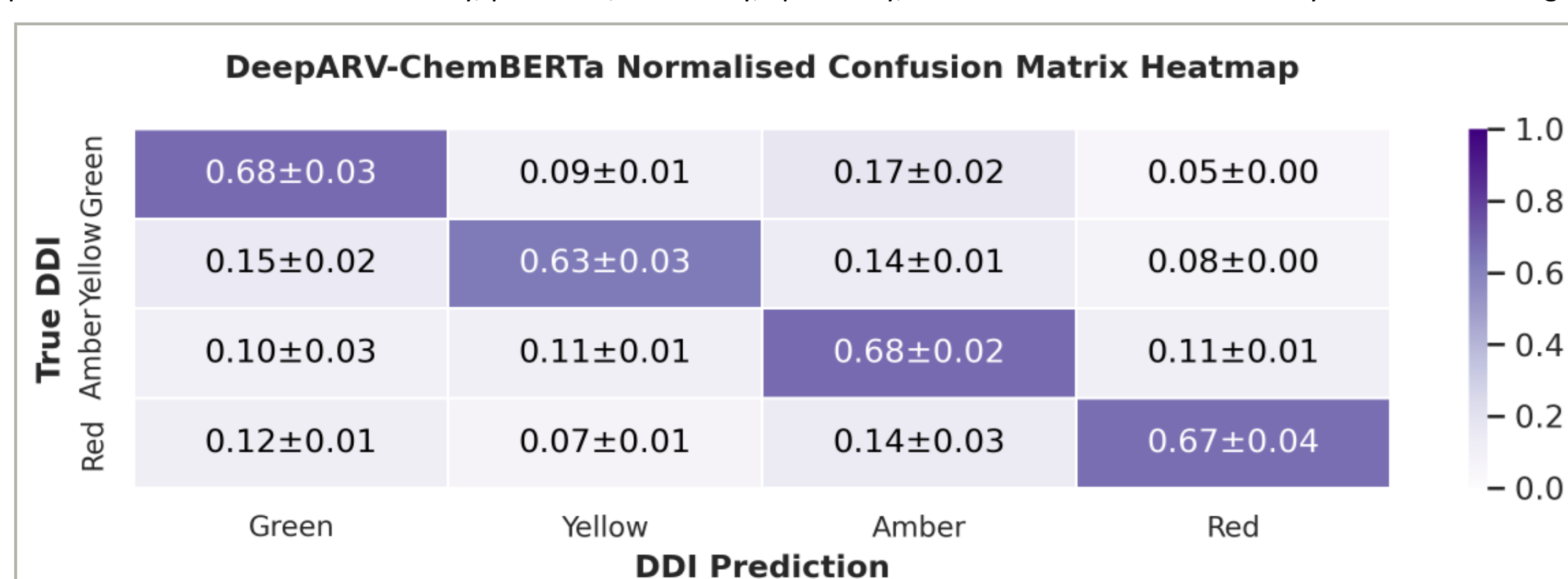


Fig2. Normalised confusion matrix of DDI prediction by DeepARV-ChemBERTa versus the ground truth DDI. Heatmap was applied to represent the concentration of values by colour depth.

Conclusion

- DeepARV-ChemBERTa predicted clinically relevant DDI between ARVs and comedications with a weighted mean accuracy of 0.837 ± 0.008 and balanced accuracy (trade-off between sensitivity and specificity) of 0.776 ± 0.011 .
- DeepARV-ChemBERTa represents opportunities for a strategical integration of AI approaches during development of novel therapies as well as the identification and design of clinical DDI studies.

References

- [1] Seden, K., Back, D. and Khoo, S., 2009. Antiretroviral drug interactions: often unrecognized, frequently unavoidable, sometimes unmanageable. *J. Antimicrob. Chemother.*, 64(1), pp.5-8. [2] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), pp.436-444. [3] Chithrananda, S., Grand, G. and Ramsundar, B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. arXiv preprint arXiv:2010.09885 (2020). [4] Liverpool HIV Drug Interaction, 2022. <https://www.hiv-druginteractions.org/>.

Acknowledgements

HIV Website



Imperial College
London