Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230080

Does Using a Stacking Ensemble Method to Combine Multiple Base Learners Within a Database Improve Model Transportability?

Cynthia YANG[a,1], Egill A. FRIDGEIRSSON[a], Jan A. KORS[a], Jenna M. REPS[a,b], Peter R. RIJNBEEK[a], Jenna WONG[c] and Ross D. WILLIAMS[a]
[a] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA*
[c] *Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA*

Abstract. We investigated a stacking ensemble method that combines multiple base learners within a database. The results on external validation across four large databases suggest a stacking ensemble could improve model transportability.

Keywords. Clinical prediction model, external validation, stacking ensemble

1. Introduction

When developing a clinical prediction model, it is impossible to know beforehand which modeling method is suitable for a particular prediction task. Additionally, prediction performance often drops when a model is transported to another database [2]. A stacking ensemble provides the opportunity to combine predictions from multiple base learners [1]. In [2], the authors show that ensembles combining lasso logistic regression models trained on different databases transported better than single database models. Our aim is to investigate whether using a stacking ensemble method to combine different base learners within a single observational health database improves model transportability.

2. Methods

We developed models using the Observational Health Data Sciences and Informatics (OHDSI) Patient-Level Prediction framework [3]. We used three large claims databases from the United States of America and one large electronic health record database from

[1] Corresponding Author: Cynthia Yang, c.yang@erasmusmc.nl. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Germany with data mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM): IBM MarketScan® Commercial Database (CCAE), IBM MarketScan® Multi-State Medicaid Database (MDCD), IBM MarketScan® Medicare Supplemental Database (MDCR), and IQVIA Disease Analyser Germany EMR (IQVIA Germany). Each site obtained institutional review board approval for the study or used de-identified data. We investigated 21 prediction tasks predicting 21 different outcomes of interest [3]: "Amongst a target population of patients with pharmaceutically treated depression, which patients will develop <the outcome> during the 1-year time interval following the start of the depression treatment?". We sampled an initial study population of 100,000 patients from each database. For each prediction task and database, we developed a stacking ensemble consisting of 3 different base learners (lasso logistic regression, random forest, and XGBoost) and a single meta-learner (logistic regression). A random subset of 75% of the patients was used for training and hyperparameter tuning of the base learners and the predictions of the base learners on the remaining 25% of the patients were used to train the meta-learner. To assess model transportability, we externally validated each model across the other three databases and evaluated the area under the receiver operating characteristic curve (AUC).

3. Results

On average, the stacking ensemble resulted in small positive AUC differences (Figure 1). All differences were significantly different from zero ($p < 0.05$), with the exceptions of lasso logistic regression in CCAE and IQVIA Germany, and XGBoost in CCAE.

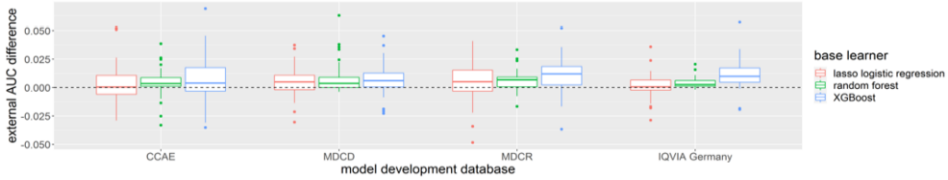

Figure 1. $External\ AUC\ difference = ensemble\ AUC - base\ learner\ AUC$. E.g., for CCAE, the blue box plot shows the ensemble AUC minus the XGBoost AUC across all outcomes and across all other databases.

4. Discussion and conclusion

The results suggest that using a stacking ensemble method can improve transportability of prediction models. However, the AUC differences were generally small, with the largest gain in AUC found for using the stacking ensemble instead of XGBoost alone. Future research may consider a larger set of base learners and different meta-learners.

References

[1] Wolpert DH. Stacked generalization. Neural Netw. 1992 Jan;5(2):241–59.
[2] Reps JM, et al. Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increasing model transportability. BMC Med Inform Decis Mak. 2022 May;22(1):1-4.
[3] Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Aug;25(8):969-75, doi: 10.1093/jamia/ocy032.