Check for updates

Commentary: Automated Machine Learning Model Development for Intracranial Aneurysm Treatment Outcome Prediction: A Feasibility Study

Markus Huber[1], Markus M. Luedi[1] and Lukas Andereggen[2,3]*

[1] Department of Anaesthesiology and Pain Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, [2] Department of Neurosurgery, Kantonsspital Aarau, Aarau, Switzerland, [3] Faculty of Medicine, University of Bern, Bern, Switzerland

A Commentary on

Automated Machine Learning Model Development for Intracranial Aneurysm Treatment Outcome Prediction: A Feasibility Study
by Ou, C., Liu, J., Qian, Y., Chong, W., Liu, D., He, X., Zhang, X., and Duan, C.-Z. (2021). Front. Neurol. 12:735142. doi: 10.3389/fneur.2021.735142

We read with great interest the article by *Ou and colleagues* (1) reporting on the application of an automated machine learning (AutoML) approach to predict recanalization after endovascular aneurysm occlusion. The authors are commended on accounting for key factors in outcome prediction such as (i) imbalanced datasets (2) by considering both the area under precision-recall curve (AUPRC) and the area under receiver-operating characteristic curve (AUROC), as well as the F1-score, (ii) the risk of overfitting by performing repeated cross-validations of the training and evaluation procedure, (iii) including graphical illustrations of the model building procedure as suggested in the literature (3) and (iv) providing code examples (4). Their results underlines the increased predictive performance of an AutoML approach compared to traditional logistic regression and a typical machine learning algorithm (Random Forest). Given the high predictive performance and the ease of using statistical software—as exemplified by the code and procedures in the Python language—the AutoML tool might provide a tool to bridge the implementation gap of such methods in medical practice (5).

From our own experience, we found the following points critical in applying ML models in outcome prediction.

While the discriminatory ability of the AutoML approach is highest among the statistical approaches in the study presented, the authors did not assess the calibration of the various algorithms. Calibration gives an estimate of how well the observed outcomes and predictions agree and are crucial in the clinical decision-making (6–8), thus we argue that an assessment of the calibration could be a further step to both evaluate and compare classical statistical methods with AutoML approaches to provide a more holistic estimate of the performance of various classifiers. As it is argued that one of the main advantages of AutoML is the possibility for non-ML experts to utilize ML models without prior know-how, we would like to point out that the application of AutoML as exemplified in the software code in Figure 4 of the paper still requires rather profound knowledge of the hyperparameters of the algorithm used in the model building pipeline—in the present application more than a dozen parameters need to be set. Thus, while the AutoML framework hides most of the parameter tuning and feature selection in a more easy-to-use

software wrapper, a certain essential knowledge of ML—such as the concept of hyperparameters and cross-validation—is still required from the user to obtain robust and unbiased results. The authors mention further drawbacks of an AutoML approach, for example in terms of the black-box problems, which could be tackled by novel interpretation techniques such as SHAP values. However, while these techniques provide information regarding the importance of individual predictors, we argue that by considering the predictive performance of an ensemble of classifiers for two performance metrics jointly provides additional valuable information to compare different algorithms (9). Thus, an illustration of the performance of various algorithms within the search for the optimal pipeline of an AutoML application might provide additional and helpful information regarding the performance and robustness of both standard statistical methods such as multivariable logistic regression and modern machine learning methods.

From a clinical perspective, recanalization and recurrences following endovascular therapy of intracranial aneurysms is not infrequently encountered. The authors indeed list the number of patients analyzed and the short-term follow-up as a study limitation. However, the short follow-up time limits its validity. Although it has been shown that coiled aneurysms that showed complete occlusion at 6 months remained stable in most cases, up to 6.5% of those aneurysm occluded completely at 6-month later showed a recanalization (10). To evaluate recurrences rates

dictating the treatment effectiveness after coiling, long-term follow-up is thus warranted (11). Although a low risk of rupture of coiled aneurysms with a follow-up period of up to 20 years have been described, larger aneurysms need to be followed for a longer time period (10, 12), as do aneurysms with residual filling after the initial treatment (13). Delayed recanalization, although rare, and the possibility of de novo aneurysm formation, however calls for continuous monitoring beyond 36 months (14).

We commend the authors on presenting an interesting and important application of a novel ML approach applicable for non-AI-experts that outperforms the commonly used statistical methods in predicting treatment outcome, as the latter is of utmost importance in any clinical practice evaluating its treatment outcomes.

AUTHOR CONTRIBUTIONS

MH, MML, and LA designed and wrote the initial draft. All authors contributed to the study design and critically revised the commentary. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

1. Ou C, Liu J, Qian Y, Chong W, Liu D, He X, et al. Automated machine learning model development for intracranial aneurysm treatment outcome prediction: a feasibility study. *Front Neurol.* (2021) 12:735142. doi: 10.3389/fneur.2021.735142

2. He H, Garcia EA. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng.* (2009) 21:1263–84. doi: 10.1109/TKDE.2008.239

3. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med.* (2020) 26:1320–4. doi: 10.1038/s41591-020-1041-y

4. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes.* (2020) 13:e006556. doi: 10.1161/CIRCOUTCOMES.120.006556

5. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med.* (2017) 376:2507–9. doi: 10.1056/NEJMp1702071

6. Cearns M, Hahn T, Clark S, Baune BT. Machine learning probability calibration for high-risk clinical decision-making. *Aust N Z J Psychiatry.* (2020) 54:123–6. doi: 10.1177/0004867419885448

7. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* (2010) 21:128–38. doi: 10.1097/EDE.0b013e3181c30fb2

8. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med.* (2019) 17:230. doi: 10.1186/s12916-019-1466-7

9. Huber M, Luedi MM, Schubert GA, Musahl C, Tortora A, Frey J, et al. Machine learning for outcome prediction in first-line surgery of prolactinomas. *Front Endocrinol.* (2022) 13:810219. doi: 10.3389/fendo.2022.810219

10. Jeon JP, Cho YD, Rhim JK, Yoo DH, Kang H-S, Kim JE, et al. Extended monitoring of coiled aneurysms completely occluded at 6-month follow-up: late recanalization rate and related risk factors. *Eur Radiol.* (2016) 26:3319–26. doi: 10.1007/s00330-015-4176-3

11. Pierot L, Barbe C, Herbreteau D, Gauvrit J-Y, Januel A-C, Bala F, et al. Immediate post-operative aneurysm occlusion after endovascular treatment of intracranial aneurysms with coiling or balloon-assisted coiling in a prospective multicenter cohort of 1189 patients: Analysis of Recanalization after Endovascular Treatment of intracranial Aneurysm (ARETA) Study. *J Neurointerv Surg.* (2021) 13:918–23. doi: 10.1136/neurintsurg-2020-017012

12. Koyanagi M, Ishii A, Imamura H, Satow T, Yoshida K, Hasegawa H, et al. Long-term outcomes of coil embolization of unruptured intracranial aneurysms. *J Neurosurg JNS.* (2018) 129:1492–8. doi: 10.3171/2017.6.JNS17174

13. McDougall CG, Johnston SC, Hetts SW, Gholkar A, Barnwell SL, Vazquez Suarez JC, et al. Five-year results of randomized bioactive versus bare metal coils in the treatment of intracranial aneurysms: the Matrix and Platinum Science (MAPS) Trial. *J Neurointerv Surg.* (2021) 13:930–4. doi: 10.1136/neurintsurg-2020-016906

14. Yeon EK, Cho YD, Yoo DH, Lee SH, Kang H-S, Kim JE, et al. Is 3 years adequate for tracking completely occluded coiled aneurysms? *J Neurosur JNS.* (2020) 133:758–64. doi: 10.3171/2019.5.JNS183651