European Radiology (2019) 29:3317–3318
https://doi.org/10.1007/s00330-018-5795-2

LETTER TO THE EDITOR

Reply to "letter to the editor: use of artificial neural networks to predict anterior communicating artery aneurysm rupture: possible methodological issues"

Jinjin Liu[1] · Yunjun Yang[1] · Bing Zhao[2]

Received: 3 September 2018 / Accepted: 24 September 2018 / Published online: 9 November 2018

Dear editor,

We thank Drs. de Jong and Aquarius for their interest in our recent article and their three experiments with the imbalanced dataset to validate the exact effects of the adaptive synthetic (ADASYN) sampling approach [1]. We also appreciate their valuable comments on the imbalanced data issue.

Their experiment results were similar to our study. The ADASYN approach improved the value of area under the curve (AUC) of the prediction. In their experiments, the overall AUC of the prediction for the imbalanced dataset and synthetically generated cases and the imbalanced dataset without ADASYN approach was 99.7% and 98.9%, respectively. In our study, the AUC for our actual dataset and synthetically generated cases and the actual dataset was 95.0% and 94.8%, respectively. Although they found the AUC for the total synthetic dataset was 85.8%, this drop still results in a relative accurate prediction in their cases. All the three AUC values showed an excellent discrimination of outcomes with the almost same discriminatory performance (AUC > 0.80).

Data imbalance is commonly encountered in biomedical applications. The actual data distribution is generally unknown and it may be unrealistic to achieve a perfect distribution match between actual data and generate synthetic data using any algorithms. Although Drs. de Jong and Aquarius proposed the use of other techniques to improve imbalanced data for use in artificial neural networks, such as node drop-outs [2] and the Matthews correlation coefficient [3], these techniques may still cannot intrinsically solve the imbalanced data issue. The node drop-outs method is a good technique to prevent overfitting issue and not used to handle imbalanced data problem. Matthews Correlation Coefficient is widely used in bioinformatics. However, as one of a number of metrics for the purpose of classifying minor class, there are still many shortcomings in existing methods to handle imbalanced data and problems yet to be addressed [4]. The best way to handle the imbalanced data issue is to collect more minority data and improve data quality, which is idealization or not easy to be realized for many reasons in a real-world data.

Unruptured intracranial aneurysms are common in the general population. There are only a small number of aneurysms rupture in the long-term follow-up. The natural history of unruptured intracranial aneurysms are unknown. Patients with ruptured and unruptured aneurysm are generally imbalanced. In our study, most of cases with intracranial aneurysms were treated after rupture and a relatively small number of patients with unruptured aneurysms were enrolled. Although we used the ADASYN approach to avoid the imbalance bias, we found the same discriminatory performance between our actual dataset and synthetically generated cases. This model may be considered to facilitate the management of unruptured anterior communicating artery aneurysms; however, we acknowledge that the prediction model still requires further validation.

✉ Yunjun Yang
yyjunjim@163.com

✉ Bing Zhao
drzhaobing@yahoo.com

[1] Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[2] Department of Neurosurgery, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200000, China

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Bing Zhao.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because this is just a response letter to discuss the imbalanced data issue in machine learning.

Ethical approval Institutional Review Board approval was not required because this is just a response letter to discuss the imbalanced data issue in machine learning.

Methodology
• performed at one institution

References

1. He H, Bai Y, Garcia EA, Li S (2008) ADASYN: Adaptive synthetic sampling approach for imbalanced learningIEEE International Joint Conference on Neural Networks, pp 1322–1328
2. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958
3. Boughorbel S, Jarray F, Elanbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS One 12:e0177678
4. Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 5:221–232