Review of: "Chronic disease treatment default prediction with random sampling optimization."

Fengpei Li[1]

1 Columbia University

Potential competing interests: No potential competing interests to declare.

Content:

The presented paper delves into the pertinent issue of class imbalance in real-world applications, particularly in healthcare systems. The authors examine the potential benefits of random over-sampling and under-sampling to address this challenge.

Strengths:

1. Relevance: The topic is highly relevant to current challenges faced in machine learning, especially in sectors like healthcare, where the cost of incorrect predictions can be significant.

2. Comprehensive Introduction: The introduction effectively sets the context by highlighting real-world scenarios (like fraud detection) where class imbalance poses a challenge.

3. Methodology: The use of a real-world healthcare dataset strengthens the paper's claims. Additionally, the use of five distinct classification algorithms for evaluation ensures a comprehensive assessment.

4. Thorough Analysis: The research thoroughly examines pre-sampling, over-sampling, and under-sampling scenarios, providing a well-rounded view of the issue.

Areas for Improvement:

1. Comparative Assessment of Techniques: While the paper does a commendable job in discussing over-sampling and under-sampling, it falls short in comparing these with more advanced techniques like SMOTE (Synthetic Minority Over-sampling Technique) or ADASYN (Adaptive Synthetic Sampling).

2. Complexity in Presentation: Tables like 2.0, 2.1, and 2.2, while packed with information, might be intimidating for readers unfamiliar with the terminologies. Simplifying or consolidating this information might make the paper more accessible.

3. Learning Curves Analysis: While the paper mentions model performance with learning curves, it doesn't delve deep into how these curves vary across models or sampling techniques. A more in-depth analysis might provide insights into model stability and convergence.

4. Insufficient Discussion on Practical Implications: While the paper provides a strong academic perspective, it could benefit from more real-world examples or case studies, particularly in the healthcare context. For instance, what could be the potential implications of misclassifying a patient based on these models?

5. Unclear Funding Declaration: The 'Declarations' section seems truncated, lacking clarity on the funding sources, which might raise questions about potential biases.

6. Lack of a Robust Conclusion: The conclusion might be strengthened with specific recommendations or best practices based on the findings. For example, in what scenarios would over-sampling be more beneficial than under-sampling?

Concluding Remarks:

The research by Owusu-Adjei et al. presents a thorough examination of random sampling techniques in addressing class imbalances in predictive modeling. While the paper scores on relevance and comprehensive methodology, it could benefit from a clearer comparison of techniques, deeper real-world context, and more streamlined presentation. Future iterations might focus on juxtaposing traditional sampling techniques against more modern methods and elucidating their practical implications in the healthcare domain.