AJNR

Reply:

P. Tiwari and A. Madabhushi

AJNR Am J Neuroradiol 2017, 38 (11) E94
doi: https://doi.org/10.3174/ajnr.A5366
http://www.ajnr.org/content/38/11/E94

REPLY:

We thank Dr Colby for his interest in our article "Computer-Extracted Texture Features to Distinguish Cerebral Radionecrosis from Recurrent Brain Tumors on Multiparametric MRI: A Feasibility Study."

We are replying to Dr Colby's specific comment, "Therefore, while it may seem impressive that the imaging-based classifier attained 91% accuracy (10/11 cases) … we would, in fact, have attained the exact same diagnostic accuracy by ignoring all the machine-learning algorithms, relying solely on our general knowledge of the base rate that tumor recurrence is more common and assigning every holdout test case to the 'recurrence' class label without looking at a single image."

While we appreciate Dr Colby's viewpoint, we respectfully disagree with his argument for at least a few reasons.

First, the test set was curated on the basis of the availability of studies that had pathologic confirmation (via multiple biopsies or surgical resections), which, in this case, happened to consist predominantly of recurrence cases. It was not compiled to be representative of the "true" distribution of tumor recurrence and necrosis cases. However, the image-based classifier during testing did not have this information a priori. Similarly, the readers were blinded to the class distribution of cases in the test set. It is certainly conceivable that readers, when reviewing cases in clinical practice, tend to factor in the a priori distribution of recurrence and necrosis cases into their diagnosis, but there is no suggestion that this information was explicitly taken into account by the readers during the course of this study.

Second, the null classifier alluded to by Dr Colby is more of an abstract, theoretic idea. It so happens that for the number of cases ($n = 11$ with 10 cancer recurrences and 1 necrosis) considered in the test set, applying the "null classifier," which involves labeling every case as cancer recurrence, would yield the same accuracy as the machine classifier. However, one could argue that if we were to create a subset of studies that had just 1 radiation necrosis case in it, the null classifier would have an accuracy of 0% and the machine classifier would have a 100% accuracy; the machine classifier in the test was able to correctly identify the necrosis case. While the discussion of the performance of the null classifier is certainly interesting from a theoretic perspective, clearly neither the human readers nor the machine classifier in our study invoked this null classifier approach.

P. Tiwari
A. Madabhushi
Department of Biomedical Engineering
Case Western Reserve University
Cleveland, Ohio

http://dx.doi.org/10.3174/ajnr.A5366