ON-LINE APPENDIX

Statistical Analysis

Training Dataset. The TCIA/TCGA–derived training dataset included imaging and demographic information from 102 patients with *IDH*-mutant LGGs. Of these 102 patients, 21 were determined via interreader consensus to be positive for the T2-FLAIR mismatch sign and 81 were either determined via interreader consensus to be negative for the T2-FLAIR mismatch sign or had no T2-FLAIR match sign information available. For the latter group of 81 patients with *IDH*-mutant LGGs, each patient had completed neuroradiologist-derived interreader consensus information for the following: 1) primary lobe: yes/no centered on frontal lobe; 2) texture: more or less than 75% of the tumor showing homogeneous signal intensity on T1WI/T2WI; 3) margins: more or less than 75% of the tumor showing sharp/circumscribed margins; 4) T2* susceptibility blooming: present or absent; 5) contrast enhancement: present or absent; 6) cysts: present or absent; 7) necrosis: present or absent; 8) maximum tumor diameter (centimeter); 9) cortical infiltration: present or absent; 10) peritumoral edema: present or absent; 11) gliomatosis: present or absent; 12) midline shift (centimeter); and 13) hydrocephalus: present or absent. Thirty-eight of the 81 patients with *IDH*-mutant LGGs had 1p/19q-codeleted *IDH*-mutant LGGs, while the remaining 43 patients had noncodeleted *IDH*-mutant LGGs.

Training Dataset: 1p/19q-Codeletion Analytic Methods. Multivariate logistic regression was applied in a 2-step analytic process to the training set data of the 81 patients with *IDH*-mutant LGGs who either were classified by interreader consensus as negative for the T2-FLAIR mismatch sign ($n = 68$) or had no T2-FLAIR mismatch sign information available ($n = 13$). In the first step of the analytic process, a full MLR model was constructed, in which the MLR response variable (Y) distinguished 1p/19q-codeleted *IDH*-mutant LGG cases ($Y = 1$) from histologically confirmed noncodeleted *IDH*-mutant LGG cases ($Y = 0$). The set of MLR predictor variables comprised the aforementioned imaging metrics and patient age. The full MLR model was constructed with the overarching goal of identifying, among the full set of MLR predictor variables, a subset of predictor variables in which each member of the subset contributes information about *IDH*-mutant LGG 1p/19q-codeletion status that is unique from the information provided by the remaining predictor variables. Identifying this subset was based on a set of type III Wald χ^2 tests, in which the per predictor variable, the type III Wald χ^2 test, served as the pivotal quantifier for testing the null hypothesis that the unique information about *IDH*-mutant LGG 1p/19q-codeleted status provided by the predictor variable is no greater than what would be expected purely by chance. Among the complete set of predictor variables, those predictor variables in which the type III Wald χ^2 test led to rejecting the null hypothesis at the $\alpha = .10$ level statistical significance were selected as the subset of predictor variables that would be used in the diagnostic classification.

In the second step of the analytical process, a reduced MLR model was constructed in which the predictor variables were those identified in the first step of the analytic process as unique predictors of the 1p/19q-codeletion status of *IDH*-mutant LGGs. The regression equation of the reduced MLR model was then used to compute the predicted probability for the *IDH*-mutant LGGs being 1p/19q-codeleted. These predicted probabilities were then used to derive a classification algorithm rule for classifying T2-FLAIR mismatch sign–negative (or T2-FLAIR mismatch sign missing) *IDH*-mutant LGG cases as either 1p/19q-codeleted or 1p/19q-noncodeleted. The predicted probability threshold for the classification rule of the algorithm was derived by identifying the 1p/19q-codeleted predicted probability threshold that produced the largest Youden J statistic (J = Diagnostic Sensitivity + Diagnostic Specificity − 1),[1] in which the J statistic captured the overall performance of a dichotomous diagnostic classification rule when false-positive and false-negative misclassifications were assumed to carry equal weight with respect to misclassification cost.

Validation Dataset. The institutional *IDH*-mutant LGG data base included imaging and demographic information from 106 patients with LGGs. Of the 106 patients with *IDH*-mutant LLGs, 16 were classified by reader A as positive for the T2-FLAIR mismatch sign, and 19 were classified by reader B as positive for the T2-FLAIR mismatch sign. Hence, the reader A validation set of T2-FLAIR mismatch sign–negative (or T2-FLAIR mismatch sign missing) *IDH*-mutant LGG cases included a total of 90 patients with *IDH*-mutant LGGs; 55 of these patients had histologically confirmed 1p/19q-codeleted *IDH*-mutant LGGs, while the remaining 35 had histologically confirmed noncodeleted *IDH*-mutant LGGs. The reader B validation set of T2-FLAIR mismatch sign–negative (or T2-FLAIR mismatch sign missing) *IDH*-mutant LGG cases included a total of 87 patients with *IDH*-mutant LGGs; 56 of these patients had histologically confirmed 1p/19q-codeleted *IDH*-mutant LGGs, while the remaining 31 had histologically confirmed noncodeleted *IDH*-mutant LGGs.

Validation Dataset: Interreader Agreement Analytic Methods. Interreader agreement with regard to reader A and reader B's texture, T2*susceptibility blooming, T2-FLAIR mismatch sign, hydrocephalus, and primary lobe classifications was evaluated via the unweighted κ statistic.

Validation Dataset: 1p/19q-Codeletion Classification Analytic Methods. The training set–reduced MLR model equation was applied to the set of reduced model predictor variables of the T2-FLAIR mismatch sign–negative (or missing, $n = 3$) *IDH*-mutant LGG cases of the validation datasets of Readers A ($n = 90$) and B ($n = 87$) to obtain patient-specific predicted probabilities for 1p/19q-codeleted *IDH*-mutant LGGs. The validation set 1p/19q-codeleted predicted probabilities of Readers A and B were subjected to the training set–derived classification algorithm rule so that each of T2-FLAIR mismatch sign–negative (or missing) *IDH*-mutant LGG cases of the validation datasets of Readers A and B could be classified as either 1p/19q codeleted or 1p/19q noncodeleted based on whether the 1p/19q-codeleted predicted probability was greater than or equal to the training dataset established classification algorithm predicted probability threshold or less than the training dataset established classification algorithm predicted probability threshold, respectively. Finally, after the 16 cases positive for the T2-FLAIR mismatch sign of reader A and the 19 cases positive for the T2-FLAIR

mismatch sign of reader B were included as true-negatives (ie, noncodeleted) in the validation 1p/19q-codeletion classification summaries of reader A and reader B, diagnostic classification accuracy was judged per reader.

Statistical Software. The statistical software package Spotfire S+, Version 8.2, was used to conduct the MLR analyses, and the pROC[2] package of R[3] was used to conduct the diagnostic classification performance analyses.

REFERENCES

1. Youden WJ. **Index for rating diagnostic tests.** *Cancer* 1950;3:32–35 CrossRef
2. Robin X, Turck N, Hainard A, et al. **pROC: an open-source package for R and S+ to analyze and compare ROC curves.** *BMC Bioinformatics* 2011;12:77 CrossRef Medline
3. R Development Core Team. *R: A Language and Environment for Statistical Computing.* Vienna; R Foundation for Statistical Computing; 2017