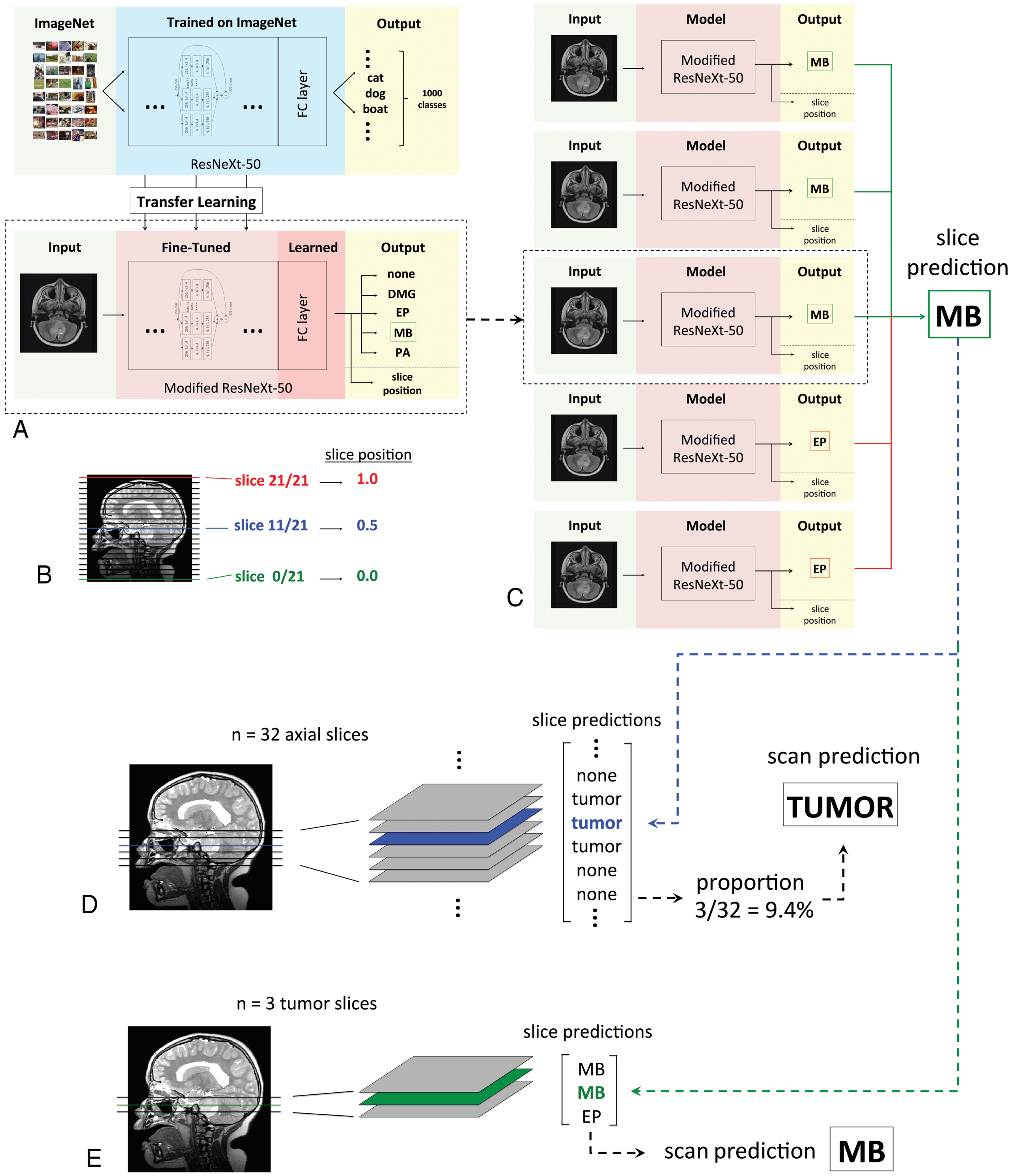
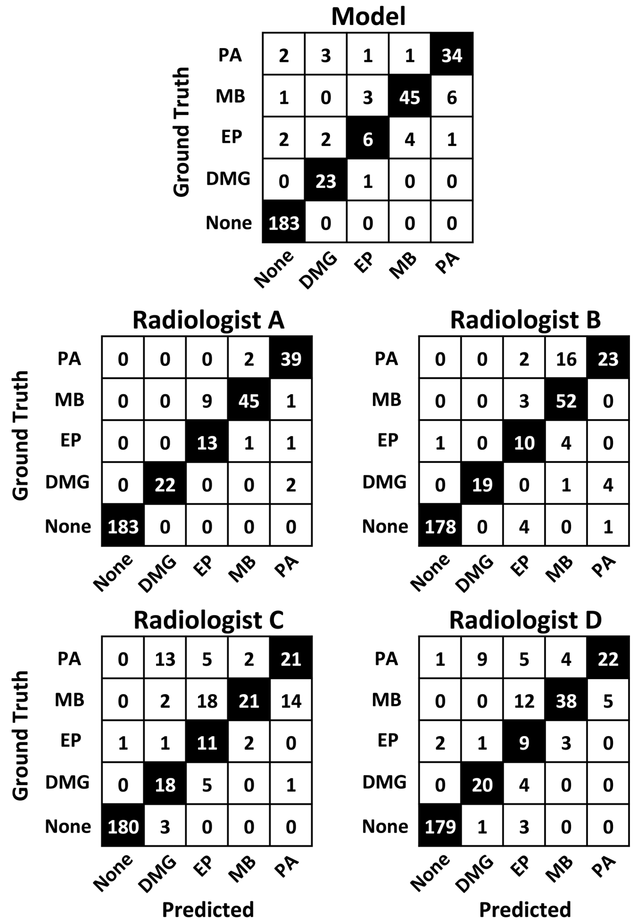


ON-LINE FIG 1. Deep learning model architecture consisting of a modified ResNext-50 pretrained on ImageNet and fine-tuned to classify individual axial slices as no tumor, MB, PF, EP, or DMG (*A*). The addition of multitask learning to predict relative slice position improves performance (*B*). The top 5 performing models are combined to create a final ensemble model for slice-level classification (*C*). Individual slice predictions are aggregated to generate scan-level predictions for tumor detection if the proportion of tumor slices exceeded a certain threshold (*D*). For scans with tumors, tumor subclass is determined on the basis of a confidence-weighted majority vote across all tumor slices (*E*).



ON-LINE FIG 2. Confusion matrices showing model and radiologists' predictions compared with ground truth.

On-line Table 1: Loss contribution of relative-slice position error on slice-level classification accuracy on validation set scans with tumors[a]

Loss Contribution	Slice-Level Accuracy	F_1 Score	False-Negative Proportion
0	0.76	0.70	0.03
10%	0.80	0.70	0.01
20%	0.72	0.70	0.01

[a] False-negative proportion indicates the proportion of scans analyzed by the model that were falsely determined to have no positive tumor slices.

On-line Table 2: Comparison of T2 and T1-T2-ADC performance on validation-set tumor classification

Sequence	F_1 (Slice-Level)	F_1 (Scan-Level)	Accuracy	False-Negative Proportion
T2	0.62	0.74	0.77	0.00
T1-T2-ADC	0.46	0.47	0.54	0.12

On-line Table 3: Model classification and detection results on the held-out test dataset

Model	Classification Accuracy	Classification F_1 Score	Detection Sensitivity	Detection Specificity	Detection AUROC
Single (top 1)	0.82	0.69	0.99	0.85	0.99
Ensemble (top 5)	0.92	0.80	0.96	1.00	0.99

Note:—AUROC indicates Area Under the Receiver Operating Characteristic curve.