Journal of Magnetic Resonance Imaging

WILEY | ISMRM ONE COMMUNITY FOR CLINICIANS AND SCIENTISTS

RESEARCH ARTICLE

Infant Brain Age Estimation With T1w/T2w Ratio MRI: A Myelination-Aware Deep Learning Approach

Hyeryn Park[1] | Young Hun Choi[2,3] | Sung-Min Gho[1]

[1]DEEPNOID Inc., Seoul, Republic of Korea | [2]Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea | [3]Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea

Correspondence: Sung-Min Gho (sungmingho@gmail.com)

Received: 22 January 2026 | **Revised:** 7 May 2026 | **Accepted:** 7 May 2026

Keywords: brain age estimation | deep learning | infant brain MRI | myelination biomarker | neurodevelopment | T1w/T2w ratio imaging

ABSTRACT

Background: Brain age estimation provides a noninvasive MRI biomarker of neurodevelopment. In infancy, rapid regionally ordered myelination reflects brain maturation, yet early-life brain age estimation remains underexplored, particularly with myelination-sensitive MRI and biologically informed modeling.

Purpose: To develop and evaluate a biologically informed deep learning framework for infant brain age estimation using T1w/T2w ratio MRI.

Study Type: Retrospective.

Population: Internal cohort: 629 infants aged 0–24 months (626 with age-appropriate myelination, train/validation/test = 376/125/125), 3 with myelin-related developmental abnormalities for qualitative review. External cohort: 10 healthy infants aged 0–15 months (5 females, 5 males).

Field Strength/Sequence: Internal: 3T; 3D gradient-echo or 2D spin-echo T1w, and 2D turbo spin-echo T2w. External: 3T; 3D gradient-echo T1w and 2D turbo spin-echo T2w.

Assessment: 3D convolutional neural networks were trained with T1w, T2w, and T1w/T2w ratio inputs using manually defined biological age labels from visual myelination assessment. The model incorporated multi-task learning for age regression, white matter segmentation, and image reconstruction.

Statistical Tests: Performance was evaluated using five-fold cross-validation with repeated random splits. Metrics included mean absolute error, root mean squared error, R^2, and Pearson and Spearman correlations. Modality differences were tested using one-way ANOVA, t-tests, and Mann–Whitney U, with Cohen's d and 95% confidence intervals. In the external cohort, absolute prediction errors were compared using the Wilcoxon signed-rank test. Statistical significance was defined as $p < 0.05$.

Results: T1w/T2w ratio models achieved the best overall performance (MAE: 1.489±0.302 months; $r = 0.966 \pm 0.012$), compared with T1w (2.055±0.944; 0.933±0.061), T2w (1.794±0.434; 0.947±0.023), T1w+T2w (1.546±0.291; 0.960±0.013), and T1w+T2w+RI (1.498±0.313; 0.963±0.012). Modality effects were significant for MAE, RMSE, R^2, r, but not for ρ ($p = 0.250$). Auxiliary-task and multi-scale modeling numerically improved performance (MAE, 1.203 months; $r = 0.979$). External validation showed the lowest error for the RI-based model (MAE, 1.16 months), and Grad-CAM highlighted myelination-relevant white matter.

Data Conclusion: T1w/T2w ratio MRI combined with biologically informed deep learning enabled accurate and interpretable infant brain age estimation. This framework showed promising cross-scanner performance and may support MRI-based assessment of early brain maturation.
Evidence Level: 3.
Technical Efficacy: 2.

1 | Introduction

Brain aging is a dynamic biological process that reflects the structural and functional maturation of the brain over time [1]. As individuals age, the brain undergoes progressive anatomical changes, particularly in gray and white matter, which are closely related to cognitive function and decline [2]. In neuroimaging, brain age estimation (BAE) has emerged as a data-driven biomarker that predicts an individual's age based on brain MRI scans [3].

By comparing predicted brain age with chronological age, BAE enables quantification of the brain age gap (BAG)–the discrepancy that may signal atypical brain development or early neuropathology [4]. A positive BAG implies accelerated aging, while a negative BAG suggests delayed maturation. Especially during early brain development, deviations from normative trajectories can have clinical implications [5].

Among the biological processes underlying brain maturation, myelination plays a central role. Myelin sheaths enable efficient neural communication and are closely associated with the development of motor, sensory, and cognitive abilities [6, 7]. During infancy, particularly within the first 2 years, myelination progresses rapidly in a regionally organized and temporally predictable pattern [8].

Given the increasing clinical use of brain MRI in infants with suspected developmental delay, there is a growing need for objective and quantitative tools to assess brain maturation. Accurate BAE during infancy may help provide an imaging-based index of age-appropriate maturation and support neuroradiologic evaluation of developmental status, while its potential clinical utility for identifying specific neurodevelopmental disorders remains to be established [4, 9].

Conventional MRI-based biomarkers, particularly T1w and T2w, have long served as indicators of brain maturation. Myelinated regions typically appear hyperintense on T1w and hypointense on T2w due to changes in lipid and water content, respectively [6, 10, 11]. These characteristic signal patterns have long been used as qualitative markers of age-appropriate white matter development. To enhance sensitivity to myelin content, the T1w/T2w ratio has been proposed as a simplified in vivo proxy for intracortical and subcortical myelination [12–15]. Initially developed for adult cortical mapping, this ratio can be readily computed from standard clinical MRI protocols and has demonstrated potential in neonatal and pediatric populations [16, 17]. Advanced imaging techniques such as mcDESPOT-derived myelin water fraction, magnetization transfer imaging, or multi-echo T2w mapping often require prolonged scan times and specialized pulse sequences [18]. They are also less feasible in pediatric practice [19]. In comparison, the T1w/T2w ratio offers a practical and accessible alternative [20, 21]. While concerns remain regarding the specificity and direct biological interpretability of the ratio, it remains a practical and scalable myelination-sensitive surrogate contrast for developmental studies and routine MRI-based research [22].

BAE using deep learning approaches has gained traction in recent years, with most models trained on convolutional neural networks (CNNs) and large-scale structural MRI datasets in adolescents, adults, and older populations [23–26]. However, early-life BAE remains underexplored because infant brain maturation is rapid, nonlinear, and highly heterogeneous, demanding age-specific modeling strategies [9]. This challenge is compounded by the limited availability of high-quality infant MRI datasets and the scarcity of large-scale external validation. In this context, deep learning models trained on the T1w/T2w ratio may capture subtle myelination-related developmental signals, enabling more biologically informed and automated estimation of brain age in infants [16, 21, 24, 27, 28].

Specifically, it remains unclear whether the T1w/T2w ratio provides added value over conventional T1w and T2w inputs for estimating MRI-derived developmental age in early life, and whether anatomically informed auxiliary tasks can further improve the biological relevance and interpretability of model predictions. In the present study, the target variable was not chronological age itself, but MRI-derived developmental age based on visual assessment of myelination patterns. Accordingly, the term brain age is used in an operational sense to denote biologically informed developmental maturation on MRI.

2 | Materials and Methods

2.1 | Ethics Approval

The internal cohort was derived from a publicly available pediatric brain MRI dataset (https://zenodo.org/record/8055666). For the independent external validation cohort, this retrospective study was conducted in accordance with the Declaration of Helsinki and approved by the institutional review board of a Seoul National University Hospital (IRB No. H-2512-021-1697). The requirement for written informed consent was waived because only de-identified routine clinical MRI data were used.

2.2 | Study Cohorts and Image Acquisition

2.2.1 | Internal Training and Test Cohort

A fully anonymized publicly available pediatric brain MRI dataset was used for internal model development. The dataset was retrospectively collected from clinical brain MRI examinations

Plain Language Summary

Assessing brain development in infants is critical for early detection of developmental delays. This study developed a deep learning model that estimates infant brain age from MRI by combining two standard scan types into a ratio image highlighting myelin, the insulating coating around nerve fibers that increases as the brain matures. Trained on 629 infants aged 0–24 months, the model predicted developmental age with a mean error of approximately 1.5 months. Attention maps confirmed the model focused on regions known to undergo early myelination. This approach showed consistent performance across two different scanners and may support objective monitoring of infant brain maturation.

retrieved from the PACS using PACS/RISCrawler and originally included patients 0–36 months of age who were scanned between January 1, 2011, and March 17, 2021 [29]. According to the published dataset description, scans transferred from outside institutions were excluded, and minor artifacts or incidental findings were retained to reflect real-world clinical heterogeneity [30]. After excluding duplicate follow-up scans and subjects older than 24 months, the final internal cohort comprised 629 infants and toddlers 0–24 months of age. Of these, 626 subjects with age-appropriate myelination were used for model development and quantitative evaluation, whereas 3 subjects identified by pediatric neuroradiologists (N.L.; 7 years of general radiology experience and 4 years of adult and pediatric neuroradiology experience) as having myelin-related developmental abnormalities were excluded from cross-validation and reserved for qualitative case review.

All MRI scans were performed on a 3T (Skyra; Siemens Healthineers, Erlangen, Germany) scanner. T1w sequences were acquired using either a 3D gradient-echo (GRE) sequence under sedation or a 2D spin-echo sequence without sedation. T2w sequences were acquired using a 2D turbo spin-echo (TSE) sequence under sedation or without sedation. Detailed acquisition parameters for each sequence are provided in Table S1.

Consistent with the source data definition, the normal cohort was defined by age-appropriate myelin maturation as assessed by board-certified pediatric neuroradiologists; minor incidental findings unrelated to myelination status were not exclusionary.

2.2.2 | External Validation Cohort

To provide an initial external assessment of model robustness when transferred to a different scanner, the trained model was evaluated on an independent institutional pediatric MRI cohort collected retrospectively at a Seoul National University Hospital and acquired on a different 3T (Premier; GE HealthCare, USA) scanner. The cohort comprised 10 subjects 0–15 months of age (5 females, 5 males). Subjects were scanned with routine pediatric brain MRI protocols (Ped MRI Brain Neonate (NICU) or Ped MRI Brain Routine), including 3D GRE T1w and 2D axial TSE T2w sequences. Acquisition parameters are summarized in Table S1. This cohort was used exclusively for inference and was not involved in training, validation, hyperparameter tuning, or model selection. Predicted developmental ages were compared with the corresponding biological ages.

2.3 | Data Labeling

For the internal cohort, developmental age labels provided with the public dataset were used as the target variable for model training. Labeling was performed on postprocessed T1w and T2w MRI using the Nora imaging platform (https://www.nora-imaging.com/). Reference charts and atlases were used to guide labeling [31, 32]. In preterm infants, corrected gestational age was calculated by subtracting the number of weeks born before 37 weeks of gestation from the age at imaging and was used only as a developmental reference to support manual labeling, not as an automatic target. The final target was a manually assigned developmental age based on MRI-visible myelination patterns.

According to the source dataset publication, 85% of cases were labeled by a pediatric radiology fellow (T.A.D.; 5 years of general and 2 years of pediatric radiology experience), with complex cases reviewed by a pediatric neuroradiologist (N.L.; 7 years of general and 4 years of adult and pediatric neuroradiology experience). The remaining 15% was independently labeled by a pediatric neuroradiologist (R.A.T.; 9 years of general and 5 years of adult and pediatric neuroradiology experience) and the pediatric radiology fellow (T.A.D.) for inter-rater reliability assessment, both blinded to clinical information and follow-up data [30].

For descriptive summarization, subjects were grouped into six developmental stages based on pediatric neurodevelopmental milestones [33]: Neonate (0–1 months), Early Infant (1–3 months), Mid Infant (3–6 months), Late Infant (6–12 months), Early Toddler (12–18 months), and Mid Toddler (18–24 months). These developmental stage categories were introduced for descriptive summarization of cohort characteristics and age-wise interpretation only. Model training used continuous developmental age labels.

2.4 | Image Preprocessing and Generation of Ratio Images

The preprocessing pipeline is illustrated in Figure 1 and consisted of four sequential steps: (1) inter-modality registration, (2) brain extraction and bias correction, (3) template-space normalization, and (4) voxel-wise RI generation.

First, T2w images were linearly registered to their corresponding T1w images using FSL's FLIRT tool [34, 35]. Second, brain extraction was applied to both modalities using FSL's BET and FAST tools to isolate brain tissue for subsequent intensity-based analyses, which also supported patient de-identification [36–38]. Bias

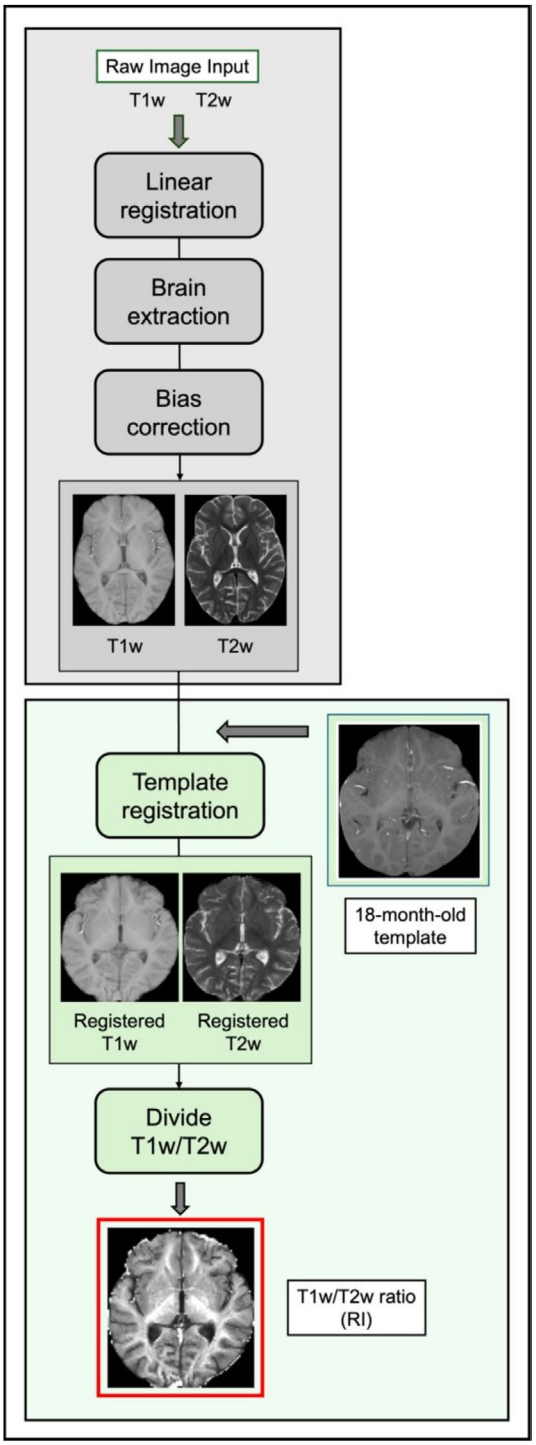

FIGURE 1 | Schematic overview of the RI construction process. Key preprocessing steps are illustrated using sample data from a representative infant. Steps performed in the native image space are marked in light gray, while those conducted in the template space are indicated in light green.

correction was then conducted in FSL, assuming that white matter voxels have uniform intensity and that low-frequency intensity variations represent image artifacts. Third, to control for age-related variations in head size, all images were affinely registered to an 18-month-old pediatric brain template obtained from a publicly available repository (https://zenodo.org/record/8055666) [30].

Fourth, the T1w/T2w ratio was generated only after registration, brain extraction, and bias correction to reduce inter-modality intensity inconsistency and low-frequency intensity variation before voxel-wise ratio computation. The final T1w/T2w ratio image (RI) was calculated by voxel-wise division of the preprocessed T1w image by the T2w image (RI = T1w/T2w) [15, 24]. To mitigate excessive skewness in the RI distribution caused by inter-modality intensity differences, outlier values above the 98th percentile—identified using kernel density estimation (KDE)—were removed to stabilize the distribution. The derived RI then served as the input to the 3D CNN model. The same preprocessing pipeline was applied to the external validation cohort.

2.5 | Model Architectures and Training

Deep learning architectures for infant BAE were evaluated using different modality combinations (T1w, T2w, T1w+T2w, RI, and T1w+T2w+RI). The study comprised two stages: (1) benchmarking both generic CNN baselines and brain-age–specific models under various modality combinations, and (2) exploring biologically informed architectural modifications. Three representative brain-age–specific architectures were selected based on their ability to directly process volumetric 3D MRI inputs without dimensionality reduction and their established use as standard baselines in previous BAE studies.

2.5.1 | Baseline and Brain Age-Specific Models

To examine the effect of input modality on predictive performance, nine representative 3D CNN architectures were evaluated, comprising six generic CNN baselines (BasicCNN, CBR-tiny [39], DenseNet [40], EfficientNet [41], ResNet variants [42], U-Net encoder [43]) and three brain-age–specific models (SFCN [25], AgeNet [26], ChenNet [9]). Each model was trained on different modality combinations to assess how T1w, T2w, and RI contribute to BAE. To ensure that the observed performance was not dependent on a single high-capacity architecture, models spanning a wide range of parameter scales were evaluated, from very lightweight networks (e.g., CBR-tiny, ~0.08 M parameters) to larger residual networks (e.g., ResNet-18, ~34 M; ResNet-34, ~64 M parameters). Detailed architectural specifications and parameter counts are provided in Tables S3 and S5.

2.5.2 | Biologically Informed Multi-Task Model With Structural Priors

The proposed model is a biologically informed multi-task 3D CNN that performs BAE alongside white matter segmentation and RI reconstruction. This architecture is designed to integrate structural knowledge of neurodevelopment—particularly myelination—into the learning process. Age regression remained the primary objective of the model, whereas the segmentation and reconstruction branches were introduced as auxiliary tasks to guide the shared encoder toward biologically meaningful representations relevant to myelination.

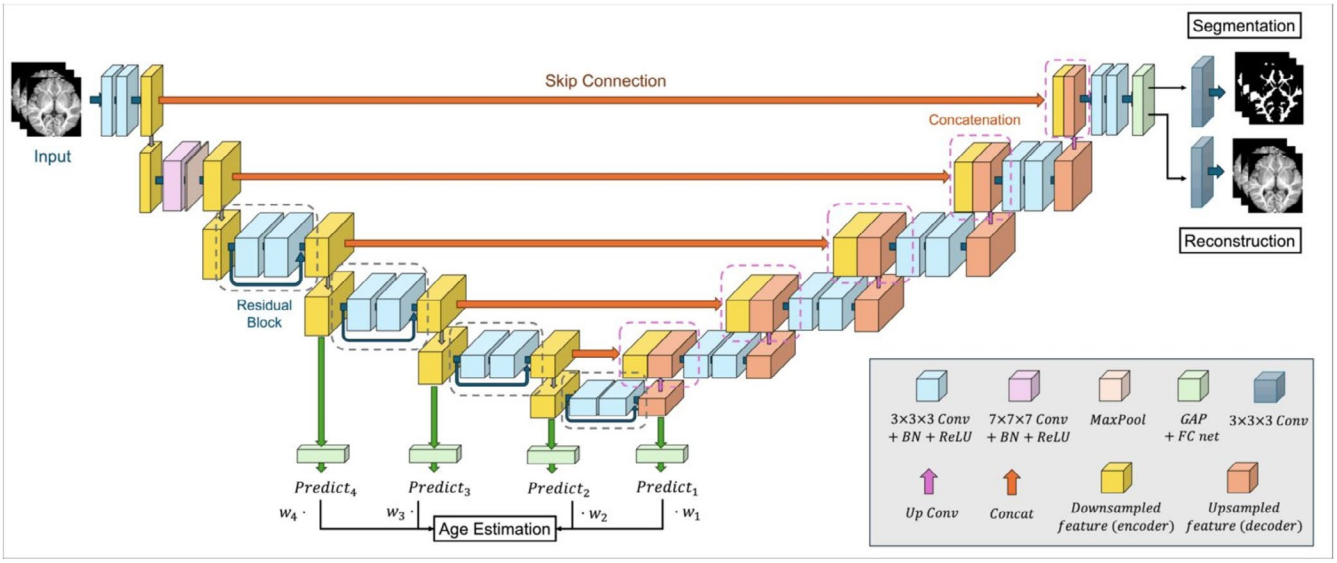

FIGURE 2 | Overview of the proposed multi-task model architecture. A shared 3D ResNet18 encoder extracts multi-scale features from the input RI. The model jointly performs brain age estimation, white matter segmentation, and RI reconstruction, using the auxiliary tasks to provide anatomical and structural regularization.

As shown in Figure 2, the model consists of a shared encoder based on a 3D ResNet18 backbone. The encoder features are extracted from multiple stages (e.g., layer3 to layer6), and the pooled feature vectors are passed to parallel shallow regressors. The intermediate age predictions from these regressors are then averaged to obtain the final age estimate. This multi-scale prediction strategy allows the model to capture both early-stage (local) and later-stage (global) developmental cues without the need for explicit deep supervision.

To incorporate biologically relevant constraints, the model includes two decoder branches. The white matter segmentation decoder predicts Infant FreeSurfer–derived white matter masks using skip connections from corresponding encoder layers, encouraging the encoder to learn features from myelination-relevant regions undergoing age-dependent structural change [44]. The reconstruction decoder reconstructs the RI input from the latent features, regularizing the encoder by preserving modality-specific signal patterns. Together, these auxiliary tasks serve as anatomical and structural priors, guiding the encoder toward neurodevelopmentally relevant representations.

2.5.2.1 | Loss Function. The model was trained with a composite loss combining three weighted components: age prediction loss ($\mathcal{L}_{age}$), white matter segmentation loss ($\mathcal{L}_{seg}$), and RI reconstruction loss ($\mathcal{L}_{recon}$), defined as $\mathcal{L}_{total} = \lambda_{age}\mathcal{L}_{age} + \lambda_{seg}\mathcal{L}_{seg} + \lambda_{recon}\mathcal{L}_{recon}$. $\mathcal{L}_{age}$ was computed as the mean squared error (MSE) between the predicted and ground-truth ages, where the prediction was obtained as a weighted sum of multi-scale predictions from intermediate encoder layers. $\mathcal{L}_{seg}$ was defined as a weighted combination of binary cross-entropy and Dice loss (α=0.5) to balance pixel-wise accuracy with spatial overlap. $\mathcal{L}_{recon}$ was computed as the MSE between the reconstructed and original RI volumes. Hyperparameters and multi-scale layer-wise weights were determined empirically through grid search to balance the three objectives [45]; the specific values used for each ablation variant are detailed in Section 2.5.3. Detailed mathematical formulations and optimization settings are provided in the Supporting Information S1.

2.5.3 | Ablation Setting and Loss Weighting

For ablation experiments, loss weights were standardized: when only one auxiliary task (Seg or Recon) was enabled, its weight was set to 1 (the other to 0) with λ_{age} fixed at 1; in weighted variants, the emphasized auxiliary objective received $\lambda = 10$. Multi-scale models used a decaying layer-wise schedule (1, 0.6, 0.3, 0.1).

2.6 | Statistical Analysis

Model performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and Pearson's (r) and Spearman's (ρ) correlation coefficients. Five-fold cross-validation was performed on the 626 typically developing infants, with each fold containing 125 test subjects and the remaining subjects split 3:1 (376/125) into training and validation sets. The procedure was repeated with two random seeds, yielding ten independent evaluations per modality. Three abnormal cases were reserved for qualitative case review.

Performance comparisons across modalities used parametric (t-test) and non-parametric (Mann–Whitney U) tests, with one-way ANOVA for group-level analyses, and effect sizes computed using Cohen's d with 95% confidence intervals. In the external validation cohort, subject-level absolute prediction errors were compared using the Wilcoxon signed-rank test with Bonferroni correction. Statistical significance was defined as $p < 0.05$, with results visualized as $-\log_{10}(p)$ values.

2.7 | Model Interpretability Analysis

For qualitative interpretation, Grad-CAM was used to visualize the spatial attention of the network [46]. For display purposes, pseudo-color images were generated by summing the available modality channels for each input configuration. Grad-CAM was then obtained for the corresponding model outputs and visualized across three convolutional layers (Conv1–3). Depending on the experimental setting, the model received single-, dual-, or multi-channel inputs.

3 | Results

3.1 | Cohort and Label Characteristics

Table 1 summarizes the distribution of cases across developmental stages, including the frequency and magnitude of prematurity correction (Prem Corr) and manual age adjustment by radiologists (Doc Adj). In the full cohort, 4.13% of subjects had corrected ages differing from their chronological age, and 3.33% had manual age adjustments differing from the recorded age. Although the average magnitude of these corrections was small (<0.3 months), both Prem Corr and Doc Adj were more frequent in the Neonate and Early Infant stages.

3.2 | Qualitative Age-Dependent Changes in T1w/T2w Ratio Images

The behavior of RI across developmental stages was examined in Figure 3. Representative axial slices from neonates (0 month), infants (2, 5, 10 months), and toddlers (15 and 22 months) are shown, highlighting both central and marginal white matter regions. Early in development, gray-white contrast was limited in both T1w and T2w images, resulting in low signal in RI. With increasing age, white matter becomes brighter on T1w and darker on T2w, and RI contrast increased, particularly in central white matter. This pattern is consistent with the known spatiotemporal pattern of myelination, where central projection fibers mature earlier than peripheral association fibers [12]. The RI therefore showed age-dependent signal changes across developmental stages.

TABLE 1 | Summary of developmental stages including sample sizes, frequencies and magnitudes of age corrections.

Stage	*N*	Prem Corr (%)	Corr Diff (mon)	Doc Adj (%)	Doc Diff (mon)
Neonate (0–1 m)	160	5.62	−0.06	2.50	−0.02
Early infant (1–3 m)	46	13.04	−0.22	13.04	−0.22
Mid infant (3–6 m)	97	1.01	−0.03	3.03	−0.10
Late infant (6–12 m)	101	6.86	−0.15	5.88	−0.14
Early toddler (12–18 m)	107	2.78	−0.05	1.85	−0.05
Mid toddler (18–24 m)	115	0.00	0.00	0.00	0.00

Abbreviations: Corr Diff = average difference from chronological age due to prematurity correction (months); Doc Adj = radiologist adjustment; Doc Diff = average difference from chronological age due to radiologist adjustment (months); Prem Corr = prematurity correction.

3.3 | Quantitative Evaluation

3.3.1 | Effect of Input Modality on Model Performance

All error-based metrics are reported in months unless otherwise specified. Table 2 summarizes model performance across input modality combinations for the evaluated architectures. In the baseline CNN models, RI showed the lowest MAE in BasicCNN (1.371±0.251), ResNet18 (1.304±0.176), ResNet34 (1.377±0.185), ResNet50 (1.324±0.150), U-Net (1.537±0.291) and ChenNet (1.306±0.208). Among the brain-age-specific models, the lowest MAE was observed with T1w+T2w+RI in SFCN (1.356±0.165) and AgeNet (1.622±0.291). Overall, RI alone showed the lowest error in several baseline models, whereas T1w+T2w+RI showed the lowest error in some brain-age–specific models.

3.3.2 | Analysis of the RI-Based Model

To further examine the RI-based setting, prediction performance was analyzed using the ResNet18 backbone, which showed the lowest MAE among the evaluated baseline CNN models under the RI input condition (1.304±0.176). As illustrated in Figure 4A, the RI-based predictions were closely aligned with the reference values, and the residuals were centered near zero with a narrower spread than those of the other input settings. This is consistent with Table 2, where RI also showed the lowest RMSE (1.842±0.246) and the highest R^2 (0.944±0.013), r (0.973±0.007), and ρ (0.968±0.006).

3.3.3 | Statistical Comparison Across Modalities

To assess the impact of input modality on model performance, mean performance averaged across all evaluated models is summarized in Table 3a, and the corresponding statistical comparisons are presented in Table 3b and Figure 4B. One-way ANOVA showed significant modality effects for MAE ($F = 18.49$), RMSE ($F = 6.73$), r ($F = 9.00$), and ρ ($F = 11.17$), whereas the modality effect for R^2 was not significant ($F = 1.27$, $p = 0.250$). At the modality level, RI showed the lowest mean MAE (1.489±0.302) and RMSE (2.103±0.369). RI and T1w+T2w+RI showed similar performance, with MAEs of 1.489±0.302 and 1.498±0.313 and RMSEs of 2.103±0.369 and 2.142±0.387, respectively. RI also showed the highest mean r (0.966±0.012) and ρ (0.961±0.014), whereas T1w+T2w showed slightly lower values (0.960±0.013 and 0.956±0.012, respectively).

Post hoc comparisons are summarized in Figure 4B(a,b), and the corresponding effect sizes are shown in Figure 4B(c).

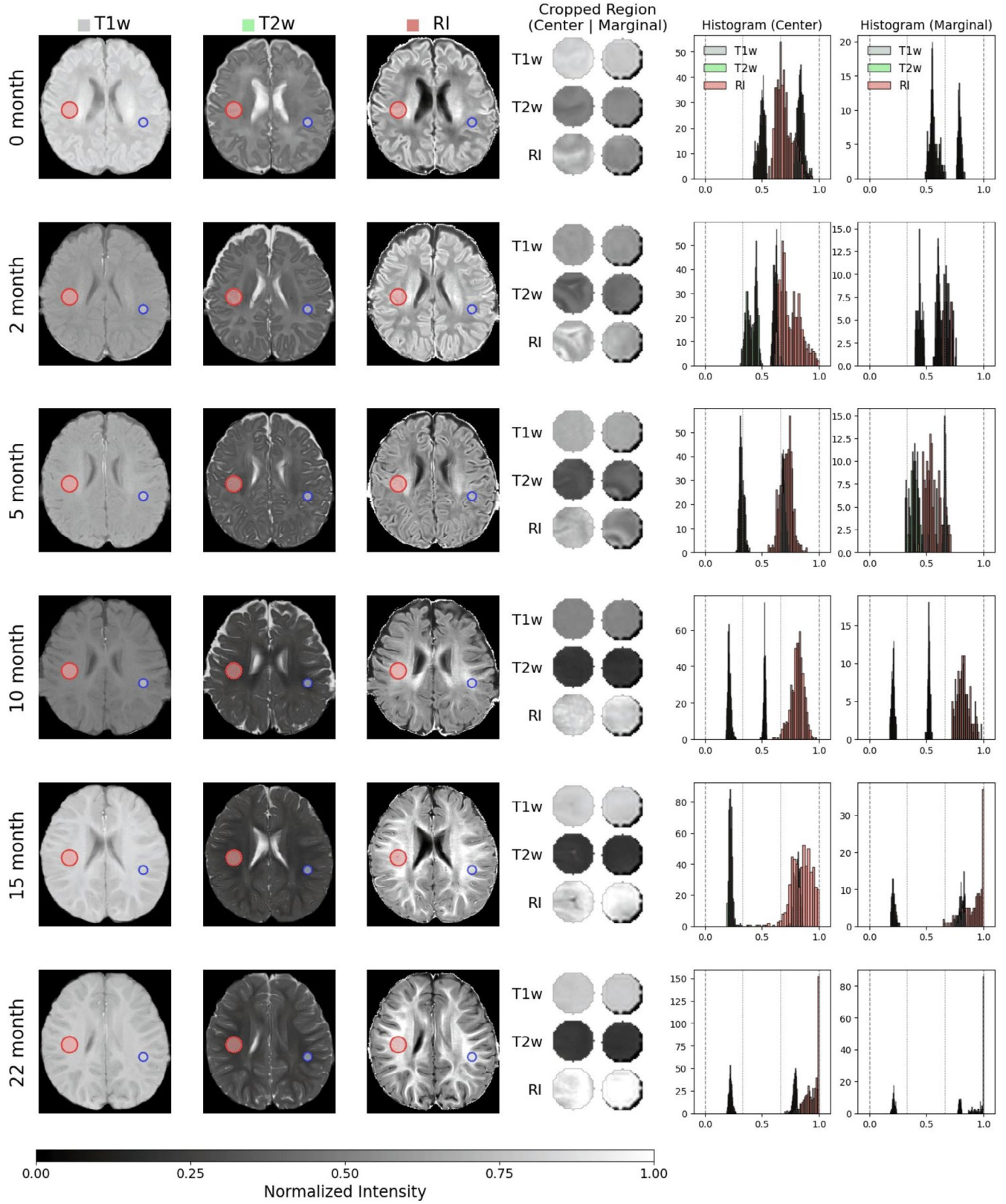

FIGURE 3 | Developmental progression in brain MR contrast across T1w, T2w, and RI modalities. Representative axial slices and Region of Interest (ROI) histograms illustrate signal evolution in central and marginal white matter from neonate to toddler stages. Signal intensities are displayed on a 0–1 normalized scale to facilitate visual comparison across modalities and regions.

Significant differences were observed between RI and T1w, T2w, and T1w+T2w across most metrics, whereas no significant difference was observed between RI and T1w+T2w+RI for MAE ($p = 0.712$), RMSE ($p = 0.344$), R^2 ($p = 0.346$), r ($p = 0.237$), or ρ ($p = 0.281$). Relative to RI, T1w and T2w showed moderate-to-large effect sizes across all metrics,

TABLE 2 | Performance comparison of deep learning–based BAE models across input modalities. Results are shown for CNN-based (top) and age-specific (bottom) architectures, reported as mean±standard deviation over ten folds.

Model	Modality	MAE ↓	RMSE ↓	R^2 ↑	r ↑	ρ ↑
CNN-based models						
BasicCNN	T1w	2.398 ± 2.327	3.375 ± 2.916	0.705 ± 0.613	0.951 ± 0.012	0.853 ± 0.300
	T2w	1.598 ± 0.202	2.363 ± 0.307	0.908 ± 0.019	0.955 ± 0.010	0.953 ± 0.008
	T1w+T2w	1.501 ± 0.193	2.305 ± 0.275	0.912 ± 0.018	0.958 ± 0.010	0.957 ± 0.007
	RI	1.371 ± 0.251	2.027 ± 0.369	0.931 ± 0.023	0.968 ± 0.011	0.964 ± 0.008
	T1w+T2w+RI	1.382 ± 0.201	2.092 ± 0.308	0.927 ± 0.018	0.964 ± 0.010	0.962 ± 0.003
CBR-tiny	T1w	1.762 ± 0.197	2.378 ± 0.248	0.905 ± 0.020	0.953 ± 0.011	0.942 ± 0.016
	T2w	1.636 ± 0.197	2.292 ± 0.290	0.912 ± 0.019	0.957 ± 0.009	0.954 ± 0.007
	T1w+T2w	1.575 ± 0.329	2.250 ± 0.470	0.914 ± 0.035	0.959 ± 0.017	0.955 ± 0.011
	RI	1.443 ± 0.146	2.002 ± 0.209	0.934 ± 0.010	0.970 ± 0.005	0.964 ± 0.005
	T1w+T2w+RI	1.408 ± 0.151	1.997 ± 0.224	0.934 ± 0.012	0.969 ± 0.006	0.964 ± 0.004
DenseNet	T1w	2.976 ± 1.277	3.898 ± 1.246	0.720 ± 0.242	0.858 ± 0.162	0.853 ± 0.186
	T2w	2.439 ± 0.359	3.316 ± 0.458	0.816 ± 0.049	0.913 ± 0.030	0.905 ± 0.030
	T1w+T2w	1.816 ± 0.330	2.492 ± 0.506	0.895 ± 0.038	0.951 ± 0.020	0.947 ± 0.018
	RI	1.882 ± 0.343	2.489 ± 0.437	0.896 ± 0.034	0.952 ± 0.018	0.950 ± 0.014
	T1w+T2w+RI	1.972 ± 0.457	2.653 ± 0.568	0.880 ± 0.050	0.946 ± 0.021	0.940 ± 0.019
EfficientNet	T1w	1.897 ± 0.353	2.730 ± 0.503	0.875 ± 0.039	0.944 ± 0.015	0.946 ± 0.012
	T2w	2.174 ± 0.480	3.048 ± 0.750	0.841 ± 0.083	0.927 ± 0.032	0.937 ± 0.016
	T1w+T2w	1.605 ± 0.237	2.300 ± 0.332	0.912 ± 0.021	0.958 ± 0.009	0.958 ± 0.007
	RI	1.603 ± 0.226	2.235 ± 0.249	0.917 ± 0.018	0.965 ± 0.008	0.963 ± 0.008
	T1w+T2w+RI	1.599 ± 0.306	2.298 ± 0.371	0.912 ± 0.027	0.959 ± 0.012	0.959 ± 0.009
ResNet18	T1w	1.694 ± 0.245	2.460 ± 0.401	0.899 ± 0.026	0.950 ± 0.014	0.948 ± 0.016
	T2w	1.531 ± 0.212	2.289 ± 0.289	0.913 ± 0.019	0.958 ± 0.009	0.957 ± 0.005
	T1w+T2w	1.462 ± 0.222	2.112 ± 0.239	0.926 ± 0.015	0.966 ± 0.006	0.962 ± 0.006
	RI	**1.304 ± 0.176**	**1.842 ± 0.246**	**0.944 ± 0.013**	**0.973 ± 0.007**	**0.968 ± 0.006**
	T1w+T2w+RI	1.356 ± 0.157	1.968 ± 0.239	0.936 ± 0.013	0.970 ± 0.006	0.967 ± 0.005
ResNet34	T1w	2.073 ± 0.516	2.963 ± 0.584	0.851 ± 0.056	0.928 ± 0.030	0.928 ± 0.026
	T2w	1.947 ± 0.411	2.741 ± 0.473	0.873 ± 0.041	0.941 ± 0.020	0.937 ± 0.019
	T1w+T2w	1.486 ± 0.253	2.108 ± 0.389	0.925 ± 0.024	0.965 ± 0.011	0.958 ± 0.010
	RI	1.377 ± 0.185	2.023 ± 0.291	0.932 ± 0.015	0.969 ± 0.005	0.966 ± 0.005
	T1w+T2w+RI	1.450 ± 0.260	2.064 ± 0.380	0.928 ± 0.024	0.968 ± 0.010	0.963 ± 0.008
ResNet50	T1w	1.747 ± 0.221	2.583 ± 0.312	0.891 ± 0.025	0.947 ± 0.011	0.946 ± 0.008
	T2w	1.541 ± 0.191	2.299 ± 0.349	0.912 ± 0.020	0.957 ± 0.010	0.953 ± 0.007
	T1w+T2w	1.408 ± 0.142	2.056 ± 0.136	0.930 ± 0.006	0.966 ± 0.004	0.962 ± 0.005
	RI	1.324 ± 0.150	1.903 ± 0.201	0.940 ± 0.010	0.972 ± 0.004	**0.968 ± 0.004**
	T1w+T2w+RI	1.337 ± 0.233	1.980 ± 0.321	0.935 ± 0.017	0.968 ± 0.008	0.964 ± 0.006

(Continues)

TABLE 2 | (Continued)

Model	Modality	MAE ↓	RMSE ↓	R^2 ↑	r ↑	ρ ↑
U-Net	T1w	1.911 ± 0.495	2.695 ± 0.608	0.872 ± 0.068	0.938 ± 0.033	0.934 ± 0.027
	T2w	1.808 ± 0.241	2.602 ± 0.330	0.888 ± 0.023	0.947 ± 0.011	0.938 ± 0.015
	T1w+T2w	1.640 ± 0.345	2.364 ± 0.439	0.905 ± 0.034	0.954 ± 0.018	0.947 ± 0.016
	RI	1.537 ± 0.291	2.176 ± 0.350	0.920 ± 0.028	0.964 ± 0.011	0.951 ± 0.021
	T1w+T2w+RI	1.645 ± 0.291	2.291 ± 0.395	0.912 ± 0.027	0.960 ± 0.009	0.946 ± 0.016
Age-specific models						
SFCN	T1w	1.657 ± 0.245	2.447 ± 0.440	0.899 ± 0.035	0.955 ± 0.014	0.955 ± 0.012
	T2w	1.401 ± 0.297	2.105 ± 0.421	0.926 ± 0.024	0.965 ± 0.012	0.963 ± 0.009
	T1w+T2w	1.390 ± 0.136	1.984 ± 0.269	0.935 ± 0.014	0.969 ± 0.007	0.966 ± 0.007
	RI	1.388 ± 0.190	2.028 ± 0.278	0.931 ± 0.017	0.969 ± 0.008	0.964 ± 0.009
	T1w+T2w+RI	1.356 ± 0.165	1.958 ± 0.242	0.936 ± 0.013	**0.971 ± 0.006**	**0.968 ± 0.004**
AgeNet	T1w	2.603 ± 0.786	3.499 ± 0.915	0.784 ± 0.124	0.895 ± 0.070	0.886 ± 0.060
	T2w	2.144 ± 0.377	2.858 ± 0.453	0.864 ± 0.036	0.934 ± 0.019	0.921 ± 0.024
	T1w+T2w	1.793 ± 0.361	2.481 ± 0.368	0.897 ± 0.029	0.951 ± 0.013	0.946 ± 0.018
	RI	1.814 ± 0.375	2.479 ± 0.442	0.896 ± 0.036	0.951 ± 0.019	0.944 ± 0.021
	T1w+T2w+RI	1.622 ± 0.291	2.231 ± 0.323	0.916 ± 0.025	0.961 ± 0.011	0.954 ± 0.008
ChenNet	T1w	1.853 ± 0.432	2.622 ± 0.557	0.883 ± 0.046	0.947 ± 0.019	0.939 ± 0.018
	T2w	1.512 ± 0.192	2.231 ± 0.296	0.918 ± 0.017	0.960 ± 0.008	0.958 ± 0.006
	T1w+T2w	1.403 ± 0.166	2.151 ± 0.238	0.924 ± 0.012	0.963 ± 0.006	0.960 ± 0.005
	RI	**1.306 ± 0.208**	**1.927 ± 0.312**	**0.938 ± 0.019**	0.970 ± 0.009	0.966 ± 0.009
	T1w+T2w+RI	1.349 ± 0.218	2.032 ± 0.300	0.932 ± 0.015	0.968 ± 0.007	0.964 ± 0.005

Note: Bold indicates the best overall, and underlined values denote the best modality per model. Arrows (↓ = low is better; ↑ = high is better) indicate the direction of better performance.

whereas the effect sizes for T1w+T2w were smaller and those for T1w+T2w+RI were negligible ($|d| = 0.05–0.16$). Although T1w+T2w showed a lower mean MAE than the single-modality inputs, its mean r and ρ remained lower than those of RI (0.960 ± 0.013 and 0.956 ± 0.012 vs. 0.966 ± 0.012 and 0.961 ± 0.014, respectively).

3.3.4 | Ablation Study on Structural and Anatomical Priors

An exploratory ablation analysis was performed using standardized loss settings and a single random seed in Table 4. Compared with baseline ResNet18 (MAE 1.285), adding either a segmentation or reconstruction decoder alone reduced MAE to 1.265 and 1.246, respectively. Combining both auxiliary decoders (+Seg +Recon) yielded MAE of 1.258 with the highest ρ among single-scale variants (0.977). Multi-scale regression further improved performance: the reconstruction-weighted multi-scale model (+Recon^b +Seg +Multi-scale) achieved the lowest MAE (1.203), lowest RMSE (1.696), highest r (0.979), and R^2 of 0.954.

3.4 | Qualitative Evaluation

3.4.1 | Model Interpretability and Regional Importance

To examine model interpretability, Grad-CAM visualizations were generated for the ResNet18 model, which was selected as the top-performing baseline model. As shown in Figure 5A, the RI-only model showed more focal activation in periventricular white matter and the posterior limb of the internal capsule than the T1w- or T2w-only models.

Figure 5B shows activation in white matter regions across the axial, coronal, and sagittal views in a representative 16-month-old subject. Figure 5C summarizes age-dependent attention patterns using overlays of Grad-CAM from the eight lowest-MAE representative cases in each age group. Younger age groups showed relatively diffuse and peripheral activation, whereas older age groups showed a greater concentration of activation in central white matter regions, including the corpus callosum and internal capsule.

Figure S1 shows similar modality-dependent attention patterns across additional representative cases from 2 to 15 months of

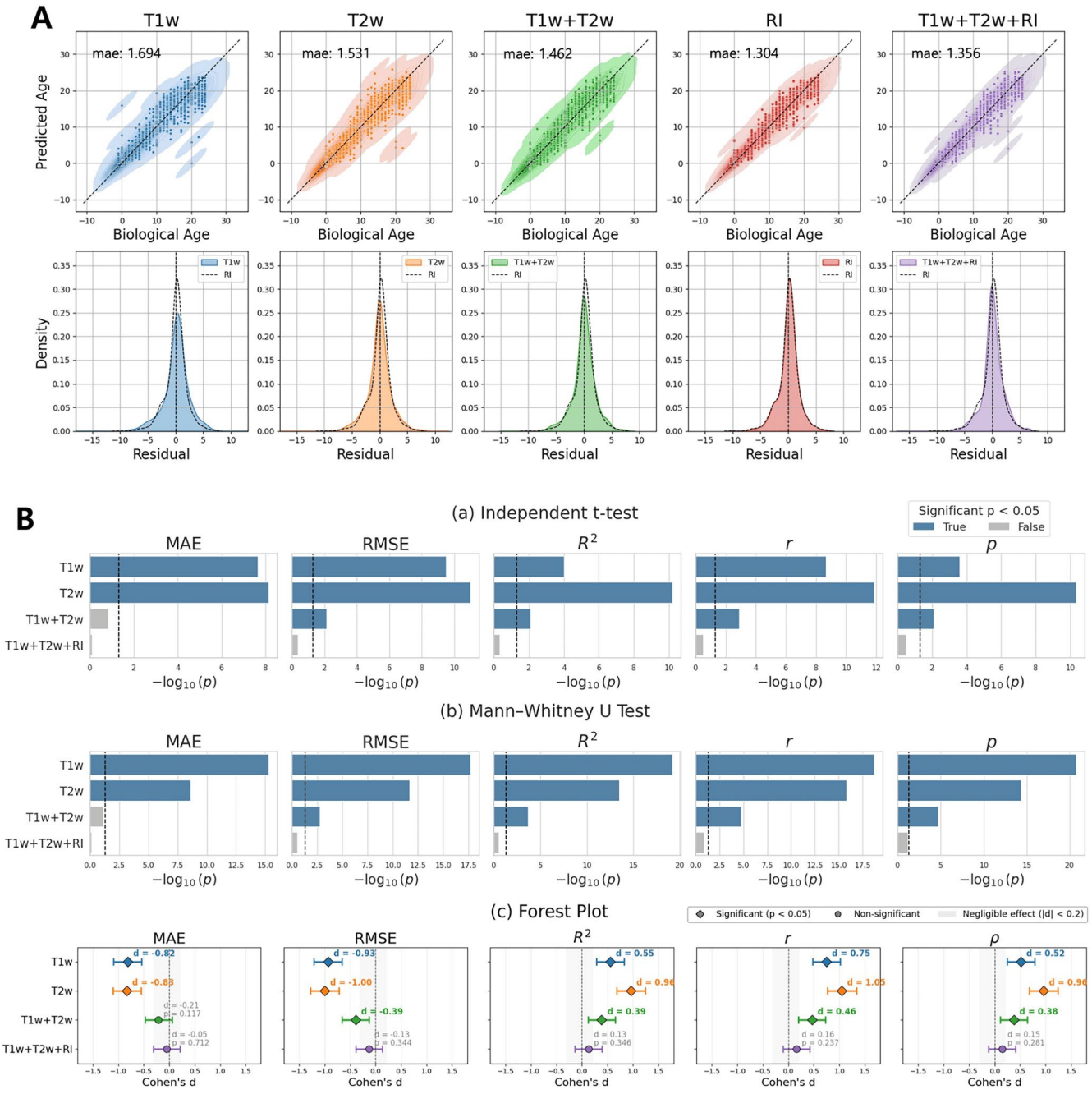

FIGURE 4 | (A) Comparison of ResNet18 performance across input modalities. Top: Ground truth vs. Predicted age. Bottom: Prediction error (Predicted—Ground truth) distributions estimated via KDE. (B) Statistical comparison of BAE performance across different MRI modalities using (a) independent *t*-test and (b) Mann–Whitney *U* test, and (c) forest plots of effect sizes (Cohen's *d* with 95% confidence intervals) for pairwise comparisons between RI-based model and the other input modalities across evaluation metrics.

age. Figure S2 provides additional multi-view examples and Figure S3 presents the individual cases underlying the age-group overlays in Figure 5C.

3.4.2 | Case Study: Myelin-Related Developmental Delay

Three subjects with myelin-related developmental abnormalities were reviewed separately to qualitatively assess clinical relevance in atypical conditions. As shown in Figure 6A, RI-based predictions were closest to the developmental ages in two of three cases: in intracerebral hemorrhage (developmental age 1 month), the RI prediction was 2.0 months versus 2.2–3.8 for other modalities; in cerebral atrophy (developmental age 12 months), T1w+T2w+RI (11.8) and RI (12.6) yielded the closest estimates, while T1w, T2w, and T1w+T2w overestimated to 14.4–14.9 months. Figure 6B shows the corresponding RI-based Grad-CAM visualizations, where activation appeared predominantly in deep and periventricular white matter regions rather

TABLE 3 | Comparison of model performance across input modalities (a) and one-way ANOVA results evaluating modality effects on each performance metric (b).

(a) Performance by input modality						(b) ANOVA by Metric		
Modality	**MAE ↓**	**RMSE ↓**	**R^2 ↑**	**r ↑**	**ρ ↑**	**Metric**	***F*-statistic**	***p***
T1w	2.055 ± 0.944	2.880 ± 1.134	0.844 ± 0.209	0.933 ± 0.061	0.921 ± 0.111	MAE	18.49	$< 10^{-29}$
T2w	1.794 ± 0.434	2.559 ± 0.547	0.888 ± 0.049	0.947 ± 0.023	0.943 ± 0.022	RMSE	6.73	$< 10^{-9}$
T1w+T2w	1.546 ± 0.291	2.238 ± 0.370	0.916 ± 0.027	0.960 ± 0.013	0.956 ± 0.012	R^2	1.27	0.250
RI	**1.489 ± 0.302**	**2.103 ± 0.369**	**0.925 ± 0.026**	**0.966 ± 0.012**	**0.961 ± 0.014**	r	9.00	$< 10^{-13}$
T1w+T2w+RI	1.498 ± 0.313	2.142 ± 0.387	0.922 ± 0.028	0.964 ± 0.012	0.959 ± 0.012	ρ	11.17	$< 10^{-17}$

Note: Bold indicates the best overall value; underline indicates the second-best.

TABLE 4 | Ablation study on the effect of auxiliary tasks and architectural variations in brain age prediction. The baseline ResNet18 (trained with a single seed) was extended with segmentation, reconstruction, and multi-scale regression tasks to evaluate their individual and combined contributions.

Model variant	MAE ↓	RMSE ↓	R^2 ↑	r ↑	ρ ↑
Baseline ResNet18	1.285	1.743	0.951	0.976	0.975
+ Segmentation Decoder	1.265	1.699	**0.954**	0.977	0.973
+Reconstruction Decoder	1.246	1.711	0.953	0.976	0.973
+Seg +Recon Decoders	1.258	1.744	0.951	0.977	**0.977**
+Seg[a] +Recon +Multi-scale	1.223	1.732	0.952	0.976	0.974
+Recon[b] +Seg +Multi-scale	**1.203**	**1.696**	**0.954**	**0.979**	0.975

Note: The bold values indicate the best performance for each metric.
[a]Segmentation-weighted ($\lambda_{seg} = 10$).
[b]Reconstruction-weighted ($\lambda_{recon} = 10$).

than at the lesion itself, consistent with the age-dependent attention trend in Figure 5C.

3.4.3 | External Validation

Figure 6C summarizes absolute prediction errors in the independent external cohort across modality settings. The RI-based model showed the lowest MAE (1.16), followed by T1w+T2w+RI (1.68), T1w+T2w (2.03), T2w (2.09), and T1w (4.77). Pairwise comparisons of subject-level absolute prediction error were performed using the Wilcoxon signed-rank test. Compared with RI, absolute prediction error was significantly higher for T1w, T2w, and T1w+T2w, whereas no significant difference was observed between RI and T1w+T2w+RI ($p = 0.322$). The T1w-based predictions showed the largest error and a marked tendency toward underestimation.

4 | Discussion

In this study, the RI showed the best overall performance for infant developmental age estimation across the evaluated input settings. In an exploratory ablation analysis performed under a standardized single-seed setting, the biologically informed multi-task and multi-scale framework yielded additional numerical improvement over the baseline. Together, these findings suggest that the RI can serve as a maturation-sensitive contrast for MRI-derived developmental age estimation and that anatomically and structurally informed auxiliary supervision may further support learning of developmentally relevant image features.

4.1 | Myelination-Aware Developmental Age Estimation

Early postnatal brain development is characterized by rapid, nonlinear, and spatially heterogeneous myelination [16, 20], with T1w/T2w-derived contrast reflecting age-related changes in major white matter structures during the first 6–9 months of life [20]. The present findings suggest that RI provides developmental information more closely aligned with myelination-sensitive contrast than T1w or T2w alone. RI alone performed similarly to the three-channel input and consistently outperformed T1w, T2w, and T1w+T2w across most metrics, suggesting that RI may provide a compact and developmentally informative representation by integrating complementary contrast from T1w and T2w images. The advantage of RI extended beyond absolute error to correlation-based metrics,

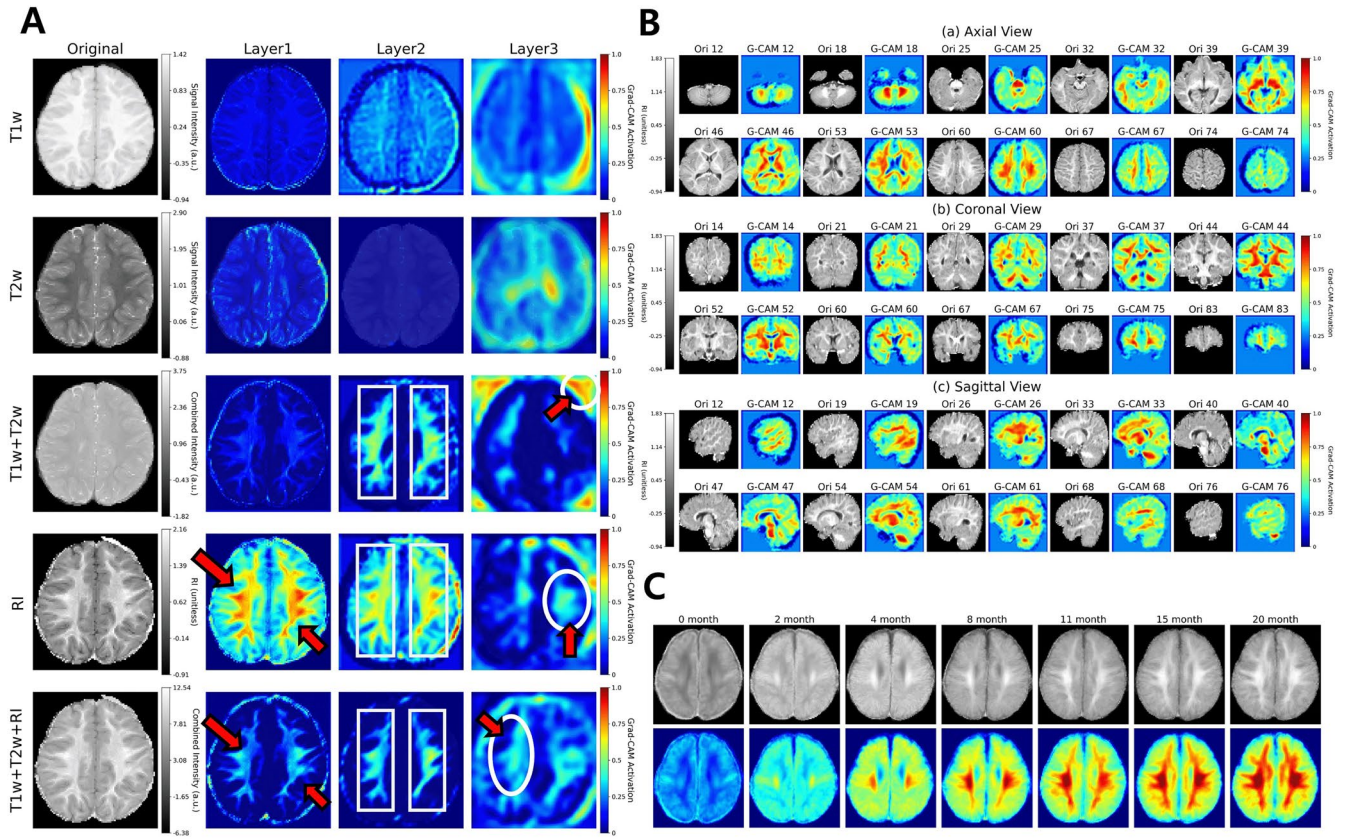

FIGURE 5 | Grad-CAM visualizations using a standard colormap (red = high attention, blue = low attention), with colorbars indicating relative activation intensity. T1w and T2w intensities are displayed as z-score normalized arbitrary units (a.u.) corresponding to model input values, whereas the RI is unitless. (A) Attention maps from three convolutional layers, highlighting modality-specific differences in network focus. (B) Representative 16-month-old case shown in (a) axial, (b) coronal, and (c) sagittal views, illustrating spatial attention across layers. (C) Developmental attention trends across representative age groups, visualized as overlays of Grad-CAM from the eight lowest-MAE representative cases for each age group. The corresponding Supporting Information figures provide additional representative cases across ages (Figure S1), randomly sampled multi-view examples (Figure S2), and the individual cases underlying the age-group overlays (Figure S3).

indicating better reflection of the relative ordering of developmental progression. The consistency of this trend across architectures with substantially different parameter scales suggests that the observed advantage may reflect a property of the input representation rather than an effect specific to model size or cohort-specific memorization.

4.2 | Interpretation in Delayed Myelination and External Validation

Case-level analysis further supports the biological plausibility of the RI-based framework. In two of three abnormal cases, RI-based predictions were closest to the adjusted developmental age, whereas the tri-modal model performed best in the remaining case. In the delayed myelination cases reviewed in this study, T1w+T2w tended to yield estimates closer to chronological age, whereas RI-based input more often produced estimates closer to developmental age; however, given the very small number of atypical cases, this observation should be interpreted cautiously. Grad-CAM visualizations also showed attention concentrated in myelination-relevant white matter regions rather than lesion-centered or nonspecific features, supporting the interpretability of the model.

The external validation results provide preliminary support for the robustness of RI-based inputs when transferred to a scanner from a different vendor. In the independent cohort acquired on a different vendor platform, the RI model yielded the lowest prediction error among the evaluated modality settings, while no significant difference was observed between RI and T1w+T2w+RI ($p = 0.322$). These findings suggest that the performance advantage of RI was not limited to the internal dataset. Nevertheless, the external cohort was small and composed only of healthy infants, and recent clinical literature has emphasized that deep learning–based myelination age models still require dedicated validation in infants with delayed myelination and in routine local imaging settings [47].

4.3 | Comparison with Prior Work

Prior studies of infant or pediatric BAE have primarily relied on conventional structural MRI inputs and general-purpose CNN architectures. Chen et al. reported accurate infant brain age prediction using T1w and T2w MRI with a task-specific 3D CNN [9]. Akinci et al. demonstrated the feasibility of early-life myelin maturation estimation using structural MRI in a clinical cohort [30]. More recently, Li et al. highlighted the

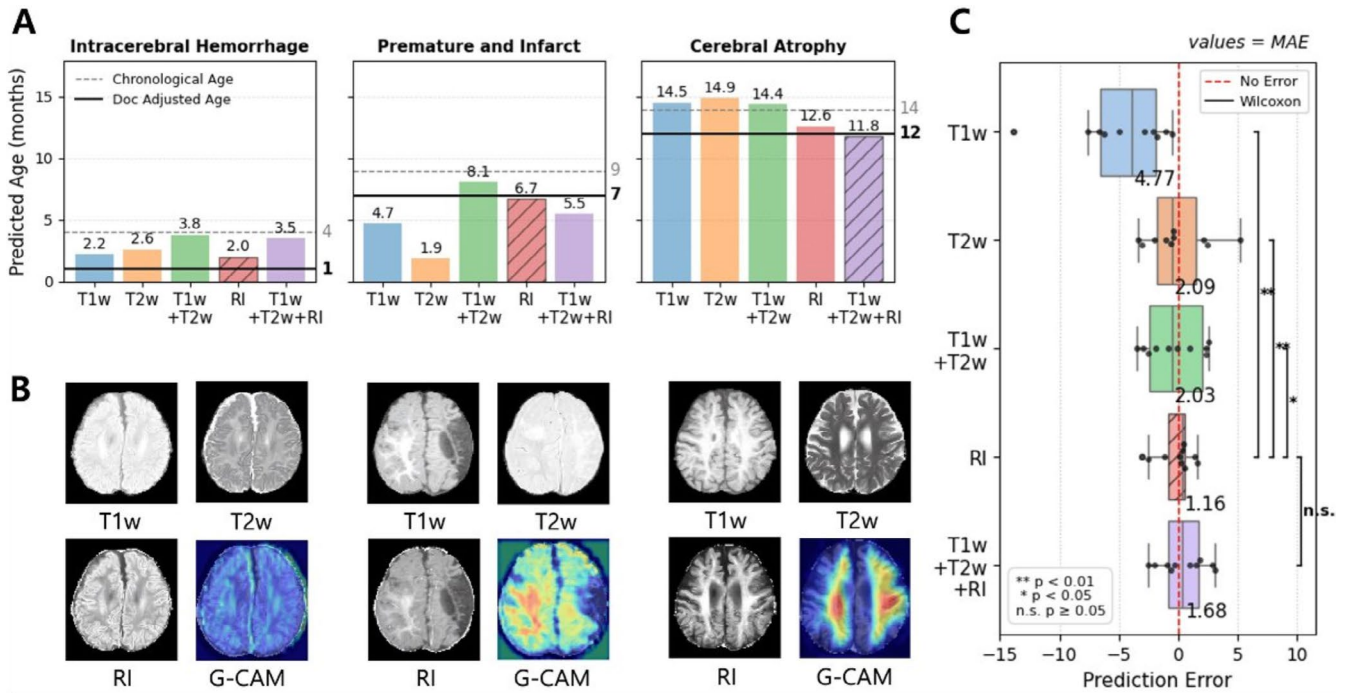

FIGURE 6 | (A) Comparison of model-predicted developmental ages across MRI modalities in three representative cases with myelin-related developmental abnormalities. (B) Representative T1w, T2w, RI, and Grad-CAM visualizations highlighting myelination-sensitive regions contributing to the prediction. (C) Distribution of prediction errors (Predicted—Ground truth) across MRI modalities in the external validation dataset. The red dashed line indicates perfect agreement between the predicted and true ages.

value of biologically informed or attention-guided modeling strategies in developmental neuroimaging [48]. Building on these studies, the present work extends the literature in several ways. First, it systematically evaluates the RI as a myelination-sensitive input contrast for infant developmental age estimation. Second, it introduces a biologically informed multi-task framework combining age prediction with white matter segmentation and RI reconstruction. Third, it compares multiple baseline and brain-age–specific architectures across several modality combinations. Finally, it supplements quantitative performance analysis with abnormal-case review, Grad-CAM visualization, and preliminary cross-vendor observation. Taken together, these contributions should be interpreted as methodological rather than definitive clinical advances; the present framework is intended to support objective MRI-based assessment of early brain maturation, not to establish diagnostic utility for specific neurodevelopmental disorders.

4.4 | Limitations

First, this was a retrospective study based on a single-center internal cohort acquired on one scanner vendor and field strength, which may have increased the risk of cohort-specific learning. The external validation cohort was small and included only healthy infants. This concern is amplified by the limited availability of large, diverse public infant MRI datasets, which constrains external benchmarking, independent replication, and fair comparison across methods. Broader multi-center validation across vendors, field strengths, and acquisition protocols, including both typically and atypically developing infants, remains necessary.

Second, the target labels were based on expert visual assessment of myelination patterns, which remains dependent on expert interpretation despite being clinically meaningful for developmental neuroimaging. In preterm infants, corrected gestational age was used only as a developmental reference and may not fully capture biological differences between extrauterine and in utero developmental trajectories [49, 50]. Inter-rater reliability was not quantitatively assessed.

Third, because the RI was derived from conventional weighted images, it may still be affected by acquisition-specific intensity scaling, prescan normalization, and residual field inhomogeneity despite preprocessing. Accordingly, RI should be interpreted as a practical myelination-sensitive surrogate contrast rather than a fully quantitative biophysical marker of myelin content.

Fourth, although the attention maps provided qualitative insight into regions contributing to prediction, they do not substitute for region-specific quantitative myelin measures and cannot by themselves establish causal biological mechanisms.

Finally, the auxiliary-task and multi-scale ablation results were exploratory: they were performed under a standardized single-seed setting and provide supportive evidence regarding anatomical and structural supervision rather than definitive evidence of superiority.

5 | Conclusion

The T1w/T2w ratio, combined with a biologically informed multi-task deep learning framework, enabled accurate and

interpretable estimation of MRI-derived developmental age in infancy. Because this approach relies on standard clinical MRI sequences, it may provide a practical and reproducible method for MRI-based assessment of early brain maturation.

Funding

This work was supported by the Technology Innovation Program (RS-2025-02221011, Development of Medical-Specialized Multimodal Hyperscale Generative AI Technology for Global Integration) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

References

1. N. Raz and K. M. Rodrigue, "Differential Aging of the Brain: Patterns, Cognitive Correlates and Modifiers," *Neuroscience & Biobehavioral Reviews* 30 (2006): 730–748.

2. J. Dobbing and J. Sands, "Quantitative Growth and Development of Human Brain," *Archives of Disease in Childhood* 48 (1973): 757–767.

3. S. He, Y. Feng, P. E. Grant, and Y. Ou, "Deep Relation Learning for Regression and Its Application to Brain Age Estimation," *IEEE Transactions on Medical Imaging* 41 (2022): 2304–2317.

4. L. Baecker, R. Garcia-Dias, S. Vieira, C. Scarpazza, and A. Mechelli, "Machine Learning for Brain Age Prediction: Introduction to Methods and Clinical Applications," *eBioMedicine* 72 (2021): 103600.

5. K. M. Welker and A. Patton, "Assessment of Normal Myelination With Magnetic Resonance Imaging," in *Seminars in Neurology*, vol. 32 (Thieme Medical Publishers, 2012), 15–28.

6. D. C. Dean, J. O'Muircheartaigh, H. Dirks, et al., "Characterizing Longitudinal White Matter Development During Early Childhood," *Brain Structure and Function* 220 (2015): 1921–1933.

7. R. D. Fields, "White Matter in Learning, Cognition and Psychiatric Disorders," *Trends in Neurosciences* 31 (2008): 361–370.

8. M. Watanabe, O. Sakai, A. Ozonoff, S. Kussman, and H. Jara, "Age-Related Apparent Diffusion Coefficient Changes in the Normal Brain," *Radiology* 266 (2013): 575–582.

9. J. V. Chen, G. Chaudhari, C. P. Hess, et al., "Deep Learning to Predict Neonatal and Infant Brain Age From Myelination on Brain MRI Scans," *Radiology* 305 (2022): 678–687.

10. A. J. Barkovich, *Pediatric Neuroimaging* (Lippincott Williams & Wilkins, 2005).

11. H. C. Kinney, B. A. Brody, A. S. Kloman, and F. H. Gilles, "Sequence of Central Nervous System Myelination in Human Infancy: II. Patterns of Myelination in Autopsied Infants," *Journal of Neuropathology and Experimental Neurology* 47 (1988): 217–234.

12. M. F. Glasser and D. C. Van Essen, "Mapping Human Cortical Areas in Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI," *Journal of Neuroscience* 31 (2011): 11597–11616.

13. M. Ganzetti, N. Wenderoth, and D. Mantini, "Whole Brain Myelin Mapping Using T1- and T2-Weighted MR Imaging Data," *Frontiers in Human Neuroscience* 8 (2014): 671.

14. H. Grydeland, K. B. Walhovd, C. K. Tamnes, L. T. Westlye, and A. M. Fjell, "Intracortical Myelin Links With Performance Variability Across the Human Lifespan: Results From T1- and T2-Weighted MRI Myelin Mapping and Diffusion Tensor Imaging," *Journal of Neuroscience* 33 (2013): 18618–18630.

15. K. Lee, M. Cherel, F. Budin, et al., "Early Postnatal Myelin Content Estimate of White Matter via T1w/T2w Ratio," in *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 9417 (SPIE, 2015), 484–490.

16. E. Filimonova, E. Amelina, A. Sazonova, B. Zaitsev, and J. Rzaev, "Assessment of Normal Myelination in Infants and Young Children Using the T1w/T2w Mapping Technique," *Frontiers in Neuroscience* 17 (2023): 1102691.

17. T. Shimozono, T. Shiiba, and K. Takano, "Radiomics Score Derived From T1-w/T2-w Ratio Image Can Predict Motor Symptom Progression in Parkinson's Disease," *European Radiology* 34 (2024): 7921–7933.

18. F. Heath, S. A. Hurley, H. Johansen-Berg, and C. Sampaio-Baptista, "Advances in Noninvasive Myelin Imaging," *Developmental Neurobiology* 78 (2018): 136–151.

19. M. Bouhrara, A. C. Rejimon, L. E. Cortina, et al., "Adult Brain Aging Investigated Using BMC-mcDESPOT–Based Myelin Water Fraction Imaging," *Neurobiology of Aging* 85 (2020): 131–139.

20. S. C. Deoni, E. Mercure, A. Blasi, et al., "Mapping Infant Brain Myelination With Magnetic Resonance Imaging," *Journal of Neuroscience* 31 (2011): 784–791.

21. J. E. Soun, M. Z. Liu, K. A. Cauley, and J. Grinband, "Evaluation of Neonatal Brain Myelination Using the T1- and T2-Weighted MRI Ratio," *Journal of Magnetic Resonance Imaging* 46 (2017): 690–696.

22. M. Arshad, J. A. Stanley, and N. Raz, "Test-Retest Reliability and Concurrent Validity of in Vivo Myelin Content Indices: Myelin Water Fraction and Calibrated T1w/T2w Image Ratio," *Human Brain Mapping* 38 (2017): 1780–1790.

23. M. Tanveer, M. A. Ganaie, I. Beheshti, et al., "Deep Learning for Brain Age Estimation: A Systematic Review," *Information Fusion* 96 (2023): 130–143.

24. J. H. Cole and K. Franke, "Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers," *Trends in Neurosciences* 40 (2017): 681–690.

25. H. Peng, W. Gong, C. F. Beckmann, A. Vedaldi, and S. M. Smith, "Accurate Brain Age Prediction With Lightweight Deep Neural Networks," *Medical Image Analysis* 68 (2021): 101871.

26. K. Armanious, S. Abdulatif, W. Shi, et al., "Age-Net: An MRI-Based Iterative Framework for Brain Biological Age Estimation," *IEEE Transactions on Medical Imaging* 40 (2021): 1778–1791.

27. S. Nerland, K. N. Jørgensen, W. Nordhøy, et al., "Multisite Reproducibility and Test-Retest Reliability of the T1w/T2w-Ratio: A Comparison of Processing Methods," *NeuroImage* 245 (2021): 118709.

28. L. B. Norbom, J. Rokicki, D. Alnæs, et al., "Maturation of Cortical Microstructure and Cognitive Development in Childhood and Adolescence: A T1w/T2w Ratio MRI Study," *Human Brain Mapping* 42 (2021): 1809–1825.

29. J. Cyriac, "PACS-RIS Crawler v1.0," 2019, https://pacs-ris-crawler.github.io.

30. T. Akinci D'Antonoli, R. A. Todea, N. Leu, et al., "Development and Evaluation of Deep Learning Models for Automated Estimation of Myelin Maturation Using Pediatric Brain MRI Scans," *Radiology: Artificial Intelligence* 5 (2023): e220292.

31. M. Staudt, I. Krägeloh-Mann, and W. Grodd, "Normal Myelination in Childhood Brains Using MRI—A Meta Analysis," *RoFo: Fortschritte Auf Dem Gebiete der Rontgenstrahlen Und der Nuklearmedizin* 172 (2000): 802–811.

32. J. A. Matsumoto, C. M. Gaskin, D. Kreitel, and S. L. Kahn, *MRI Atlas of Pediatric Brain Maturation and Anatomy* (Oxford University Press, 2015).

33. K. D. Pickrell, "Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health," *Hospitals & Health Networks* 77 (2003): 70.

34. M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images," *NeuroImage* 17 (2002): 825–841.

35. M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *NeuroImage* 62 (2012): 782–790.

36. S. M. Smith, "Fast Robust Automated Brain Extraction," *Human Brain Mapping* 17 (2002): 143–155.

37. Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Transactions on Medical Imaging* 20 (2001): 45–57.

38. Canadian Association of Radiologists Artificial Intelligence Working Group, "Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology," *Canadian Association of Radiologists Journal* 70 (2019): 107–118.

39. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," *Advances in Neural Information Processing Systems* 32 (2019): 3347–3357.

40. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), 4700–4708.

41. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *International Conference on Machine Learning* (PMLR, 2019), 6105–6114.

42. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), 770–778.

43. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* (Springer, 2015), 234–241.

44. L. Zöllei, J. E. Iglesias, Y. Ou, P. E. Grant, and B. Fischl, "Infant FreeSurfer: An Automated Segmentation and Surface Extraction Pipeline for T1-Weighted Neuroimaging Data of Infants 0–2 Years," *NeuroImage* 218 (2020): 116946.

45. P. Lerman, "Fitting Segmented Regression Models by Grid Search," *Journal of the Royal Statistical Society: Series C: Applied Statistics* 29 (1980): 77–84.

46. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2017), 618–626.

47. I. Harting, S. F. Garbade, S. D. Roosendaal, et al., "Age-Appropriate or Delayed Myelination? Scoring Myelination in Routine Clinical MRI," *European Journal of Paediatric Neurology* 52 (2024): 59–66.

48. M. Li, J. Liu, M. Yang, et al., "Myelination-Attention-Empowered Deep Learning Model Improved Brain Age Prediction in Children Below 2 Years of Age," *Pediatric Radiology* 56 (2026): 660–670.

49. T. E. Inder, L. S. de Vries, D. M. Ferriero, et al., "Neuroimaging of the Preterm Brain: Review and Recommendations," *Journal of Pediatrics* 237 (2021): 276–287.

50. V. Schmidbauer, G. Dovjak, G. Geisl, et al., "Impact of Prematurity on the Tissue Properties of the Neonatal Brain Stem: A Quantitative MR Approach," *American Journal of Neuroradiology* 42 (2021): 581–589.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** MRI Acquisition Parameters for the Internal and External Cohorts. **Table S2:** General Training Configuration for Baseline Models. **Table S3:** Architecture Summary of Baseline Models. **Table S4:** Hyperparameter Search Space for Auxiliary Losses in the Multitask Model. **Table S5:** Parameter counts of evaluated CNN architectures. **Figure S1:** Additional representative Grad-CAM examples from 2 to 15 months across the five input settings (T1w, T2w, T1w+T2w, RI, and T1w+T2w+RI). For each age, one representative case is shown together with Grad-CAM from three convolutional layers. Across these age-matched examples, the RI-only model showed more focal attention in periventricular white matter and the posterior limb of the internal capsule than the T1w- and T2w-only models. The T1w+T2w model showed more localized activation than the single-modality inputs, and the T1w+T2w+RI model showed a spatial pattern similar to that of the RI-only model. These examples are provided to complement the representative cases shown in Figure 5A. **Figure S2:** Additional representative RI and Grad-CAM examples across the examined age range. For each age group, representative cases were randomly sampled and shown individually in axial, coronal, and sagittal views using the RI image and the corresponding Grad-CAM. These supplementary examples are provided to show additional cases corresponding to the qualitative patterns illustrated in Figure 5B. **Figure S3:** Individual RI and Grad-CAM for the eight lowest-MAE cases in each age group. For each age group, the eight cases with the lowest prediction error (MAE) are shown individually using the RI image (top) and the corresponding Grad-CAM (bottom). These examples are provided to show the individual cases underlying the age-group overlays in Figure 5C.