Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36134

AI for Preterm Brain Injury Imaging: From Automated Analysis to Clinical Translation

Boyu Yuan, Weichuan Zhang*

Image Computing Laboratory, Shaanxi University of Science and Technology, Xi'an, China
**Correspondence Author. Email: zwc2003@163.com*

Abstract. Preterm brain injury is a major cause of lifelong motor, cognitive, and behavioral disability, yet its imaging phenotype changes rapidly with development and differs across cranial ultrasound (cUS) and magnetic resonance imaging (MRI). Serial cUS is suited to bedside surveillance of hemorrhage and ventricular enlargement, whereas term-equivalent structural and diffusion MRI provide more detailed assessment of tissue injury, maturation, and white matter organization. Artificial intelligence (AI) can support image-quality assessment, lesion detection, segmentation, quantitative phenotyping, developmental-age estimation, and outcome prediction, but evidence for clinical translation remains uneven. This focused review maps imaging modalities and computational tasks to neonatal decisions and synthesizes studies published from 2005 through 2025. Structural MRI segmentation has the most consistent technical evidence; cUS automation, subtle-lesion detection, longitudinal multimodal prediction, and prospective clinical use remain less mature. Dataset scale alone is not sufficient because developmental stage, acquisition differences, nonrandom missing examinations, and delayed follow-up can influence apparent performance. We therefore distinguish model accuracy from deployability and propose a minimum pipeline comprising input governance, infant-level external evaluation, calibrated uncertainty, safety-based abstention, human review, and post-deployment monitoring. Future progress depends on longitudinal multicenter cohorts, development-aware representation learning, and prospective studies that measure reporting efficiency, management changes, and safety. AI should be judged by reproducible clinical benefit rather than internal benchmark accuracy.

Keywords: Preterm brain injury; Neonatal neuroimaging; Cranial ultrasound; Magnetic resonance imaging; Artificial intelligence

1. Introduction

Improved survival after very preterm birth has increased the number of infants living with neurological risk. Intraventricular hemorrhage (IVH), periventricular leukomalacia (PVL), diffuse white matter injury (WMI), and secondary ventricular enlargement differ in onset, visibility, and prognostic meaning [1, 2]. Their consequences may include cerebral palsy, impaired cognition, and behavioral difficulty, yet acute signs can be nonspecific. Clinicians must therefore detect urgent abnormalities, grade severity, quantify tissue injury, and communicate risk while the infant may still be physiologically unstable.

DOI: 10.54254/2755-2721/2026.36134

Imaging interpretation is complicated by rapid change between approximately 24 and 40 weeks of gestational age. Cortical folding, myelination, tissue contrast, ventricular shape, and white matter organization evolve even without disease [3, 4]. A feature that is normal at one developmental stage may resemble pathology at another. Serial cranial ultrasound (cUS) supports early bedside surveillance, whereas term-equivalent magnetic resonance imaging (MRI) provides detailed structural and microstructural assessment and is associated with later development [5]. As shown in Figure 1, the value of each modality depends on timing, clinical stability, and the decision to be made. Neither modality alone captures the full trajectory from acute injury to altered maturation.

Conventional assessment also depends on expert visual interpretation and semi-quantitative grading. These approaches are clinically established but can vary between readers and may not capture diffuse abnormalities or subtle developmental patterns consistently. Artificial intelligence (AI) offers automated quality assessment, lesion detection, segmentation, quantitative phenotyping, and outcome prediction. However, high internal accuracy does not establish that a model will work across gestational ages, scanners, operators, or neonatal units. A clinically useful output must answer a defined question, arrive within the relevant workflow, and remain reviewable. This focused review therefore maps imaging modalities to clinical tasks, summarizes the strength of current AI evidence, and defines the minimum pathway from automated analysis to accountable clinical translation.

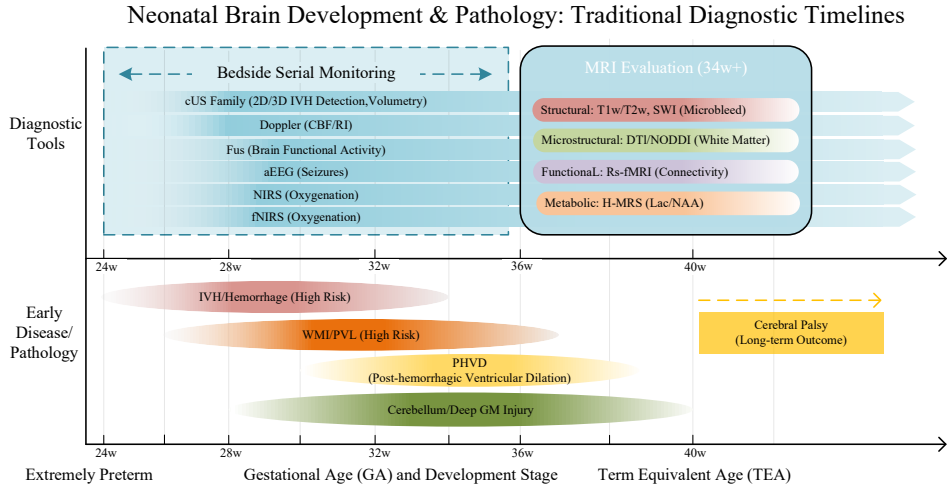

Figure 1. Clinical timing of imaging for preterm brain injury. Serial bedside methods prioritize early surveillance, whereas multimodal MRI provides more comprehensive assessment of maturation and injury near term-equivalent age. Exact schedules depend on clinical stability and local practice.

2. Review methodology

PubMed/MEDLINE, Web of Science, IEEE Xplore, and Scopus were searched for English-language studies published from January 2005 through December 2025. Search concepts combined preterm brain injury, neonatal neuroimaging, and automated analysis. After deduplication, titles, abstracts, eligible full texts, and reference lists were screened; 150 studies were retained. Eligible reports addressed segmentation, detection, grading, brain-age estimation, or neurodevelopmental prediction in fetal or neonatal populations. Adult-only, animal, descriptive, review, and methodologically incomplete reports were excluded. Because cohorts, modalities, and endpoints were heterogeneous, evidence was synthesized narratively by clinical task, model family, and validation strength rather than pooled statistically.

DOI: 10.54254/2755-2721/2026.36134

3. Imaging modalities and clinical tasks

3.1. Bedside surveillance with cranial ultrasound

cUS is portable, radiation-free, and repeatable during physiological instability, making it central to serial surveillance in the neonatal intensive care unit. It can identify moderate-to-severe IVH, ventricular enlargement, cystic PVL, and major structural abnormalities soon after birth and can track whether hemorrhage or ventricular dilatation is progressing [6]. The immediate clinical questions are usually whether an examination is adequate, whether an urgent abnormality is present, and whether its severity has changed since the previous scan.

AI tasks should follow this workflow. View and quality recognition can first confirm that required planes are present; measurement tools can quantify ventricular dimensions; detection and segmentation can localize hemorrhage or cystic change; and grading systems can prioritize examinations for specialist review. Training on isolated, high-quality frames is insufficient because routine studies contain acoustic shadowing, incomplete windows, variable gain, and operator-dependent probe orientation. Video-level evaluation is preferable when the intended input is a full examination. cUS is also less sensitive to diffuse WMI, so a negative automated result cannot exclude injury or replace later MRI [7]. Its most credible near-term role is therefore quality-aware triage and serial measurement rather than autonomous diagnosis.

3.2. Anatomical and microstructural assessment with MRI

MRI provides superior tissue contrast and multiparametric information at term-equivalent age. T1- and T2-weighted sequences support tissue segmentation, volumetry, cortical morphometry, ventricular measurement, and assessment of lesion burden [8]. Diffusion-weighted imaging can reveal acute restriction, while diffusion tensor imaging characterizes tract organization through measures such as fractional anisotropy. Functional and metabolic sequences add information about networks and tissue biochemistry [9, 10]. These inputs extend automated analysis from anatomical segmentation to punctate-lesion detection, maturation or brain-age estimation, and neurodevelopmental prediction.

The same richness creates modeling challenges. Gray-white contrast changes rapidly during the preterm period, so a fixed tissue appearance cannot be assumed across gestational ages. Motion may remove precisely the examinations from the least stable infants, and registration errors can propagate into regional measurements or multimodal fusion. Scanner field strength, sequence parameters, and reconstruction methods also alter contrast. Evaluation should therefore report performance by developmental stage and acquisition setting, and it should distinguish failure of image quality from absence of disease. MRI automation is currently most defensible for reproducible structural quantification; more complex prognostic outputs require evidence that imaging adds information beyond routinely available clinical variables.

3.3. Matching modality to decision

The clinical decision determines the acceptable error and the appropriate unit of evaluation. Screening prioritizes sensitivity, rapid turnaround, and false-negative analysis; grading requires reproducible definitions and agreement with specialists; segmentation requires anatomically plausible boundaries and stable measurements; prognosis requires calibration, clinically meaningful horizons, and improvement over established risk factors. A single headline metric cannot represent all four tasks.

Early cUS and later MRI are complementary rather than interchangeable. Longitudinal models should align observations by infant and developmental age, preserve the order of serial examinations,

DOI: 10.54254/2755-2721/2026.36134

and distinguish an unavailable MRI from a negative MRI finding. Fusion models should also be compared with cUS-only, MRI-only, and clinical baselines to show that added complexity provides added value [11]. Repeated scans from one infant must remain within the same data partition to prevent leakage. This decision-centered design reduces the risk that a model learns acquisition timing, scanner identity, or follow-up availability instead of injury biology.

The handoff between modalities should also be defined clinically. Serial cUS can establish the onset and progression of hemorrhage or ventricular enlargement and identify infants who require additional assessment. MRI can then refine the extent of non-cystic WMI, cortical or deep-gray-matter involvement, and altered tract development near term-equivalent age. A longitudinal AI system should preserve this asymmetry instead of treating all inputs as simultaneous measurements. Its output should indicate which modality supports each conclusion and whether a change represents new injury, resolution, or delayed maturation. This structure makes multimodal analysis interpretable as a care pathway rather than as an opaque fusion of features.

4. AI methods and current evidence

4.1. From anatomical priors to task-specific deep learning

Early neonatal pipelines addressed limited contrast with atlases, deformable models, spatial priors, and explicit partial-volume estimation. AdaPT, for example, used an expectation-maximization framework to preserve anatomical consistency in preterm MRI [12]. These approaches were interpretable and data-efficient, but their performance depended on registration quality and on how well population priors represented an individual infant.

CNNs shifted the field toward end-to-end feature learning. Multimodal networks learned complementary information from T1, T2, and diffusion-derived inputs, improving tissue segmentation during periods of poor gray-white contrast [13]. Residual and three-dimensional architectures then expanded the receptive field while retaining fine spatial detail [14]. Task-specific designs followed: residual U-Nets quantified diffuse white matter abnormality, graph networks represented connectome topology, and generative methods emphasized small lesions that occupy only a small fraction of the image [15–17].

Table 1. Representative computational approaches in infant brain analysis. Reported results are not directly comparable because datasets and evaluation protocols differ.

Model	Year	Architecture	Primary task	Main contribution	Reported evidence
AdaPT [12]	2013	Statistical EM	Tissue segmentation	Spatial priors and partial volume	Better than standard EM
Deep CNN [13]	2015	3D CNN	Isointense segmentation	T1/T2/FA fusion	Better than SVM/RF baselines
BrainNetCNN [18]	2017	Graph CNN	Age prediction	Edge-to-edge connectome filters	Error approximately 2 weeks
VoxResNet [14]	2018	Residual CNN	Brain segmentation	Deep residual learning	Challenge-leading result
3D ResU-Net [15]	2021	Residual U-Net	DWMA quantification	Automated lesion volumetry	Dice 0.907
BC-GCN [16]	2022	Graph CNN	Age prediction	Connectome path convolution	MAE 49.9 days
Generative detector [17]	2023	Generative model	PWML detection	Counterfactual learning	Improved small-lesion detection
SCL-GCN [19]	2024	Contrastive GCN	Outcome prediction	Supervised contrastive learning	AUC 0.72–0.75
SegMamba [20]	2024	State-space model	3D segmentation	Efficient long-range modeling	Lower memory requirement

DOI: 10.54254/2755-2721/2026.36134

Recent contrastive and state-space models aim to use limited labels more efficiently and capture long-range dependencies in volumetric data [19, 20]. Their novelty does not by itself establish clinical superiority. Table 1 shows the transition from anatomical priors to task-specific deep learning; reported values remain dataset-specific and should not be interpreted as a direct ranking.

4.2. Prediction, multimodal representation, and emerging models

Segmentation produces measurements that can be checked directly against anatomy, whereas prediction asks whether imaging contains information about maturation or later outcome. BrainNetCNN introduced filters that operate on connectome edges and demonstrated that network structure could support developmental-age estimation [18]. Later graph and supervised-contrastive approaches combined structural or functional connectivity with representation learning for brain-age and neurodevelopmental prediction [16, 19]. These studies broaden the target beyond visible lesions, but their endpoints, follow-up ages, and clinical covariates vary substantially.

Multimodal models may combine morphometry, lesion burden, diffusion measures, functional connectivity, and perinatal variables. Fusion can occur at the input, feature, temporal, or decision level. A useful model must remain evaluable when a modality is missing, because unstable infants may not receive MRI and follow-up loss is rarely random. Performance should therefore be compared with clinical-only and single-modality baselines, followed by ablation analysis showing which source contributes information.

The reviewed datasets fall into three functional groups. Small challenges such as iSeg-2017 and NeoBrainS12 provide standardized comparisons for tissue segmentation, but their limited training sets are better suited to benchmarking than to estimating clinical generalization. High-quality fetal and neonatal resources such as the developing Human Connectome Project support atlas construction, representation learning, and method development. Adult benchmarks, including BraTS or MRBrainS, may test an architecture or provide pretraining data, but performance on them does not validate a neonatal application.

Clinical cohorts address more specific questions. CINEPS links structural and diffusion MRI with neurodevelopmental prediction, while Multi-CP provides multicenter conventional MRI for cerebral-palsy prediction [19, 21]. The DWMA cohort supports lesion quantification with separate development and test infants [15]. Other studies use smaller prospective MRI cohorts or large collections of ultrasound images for outcome prediction and screening. These scales are not directly comparable: an image count may contain many correlated frames from relatively few infants, and a large research dataset may contain few relevant injuries. Final validation must therefore report the number of infants, centers, examinations, and outcome-complete cases and must use the intended population, acquisition pathway, and endpoint.

4.3. What the evidence currently supports

As shown in Figure 2, structural MRI segmentation has the most consistent technical evidence because it benefits from established anatomical labels, public challenges, and directly interpretable overlap measures. Evidence is less mature for cUS automation, punctate or diffuse lesion detection, functional imaging, multimodal prognosis, and testing across centers. This pattern reflects both data availability and the increasing distance between an image label and a clinical outcome.

Evaluation should match the intended decision. Segmentation studies should report overlap together with surface or regional errors, because a high whole-brain Dice score can conceal clinically important boundary failures. Detection studies need lesion-level sensitivity, false positives per examination, and

DOI: 10.54254/2755-2721/2026.36134

performance on negative studies rather than accuracy dominated by normal pixels. Prognostic studies should report discrimination, calibration, confidence intervals, and added value beyond a prespecified clinical baseline. Where repeated examinations exist, all images from one infant must remain in one partition.

Evidence should also be stratified by gestational age, site, scanner or ultrasound device, image quality, and injury subtype. External testing is strongest when thresholds and preprocessing are fixed before the new center is evaluated. Reader comparisons are useful for grading or detection, but agreement with a reference label is not equivalent to improved care. The present literature therefore supports automated structural quantification more strongly than autonomous diagnosis or prognosis. The central scientific gap is reliable performance across realistic settings, not the ability to fit a selected neonatal dataset.

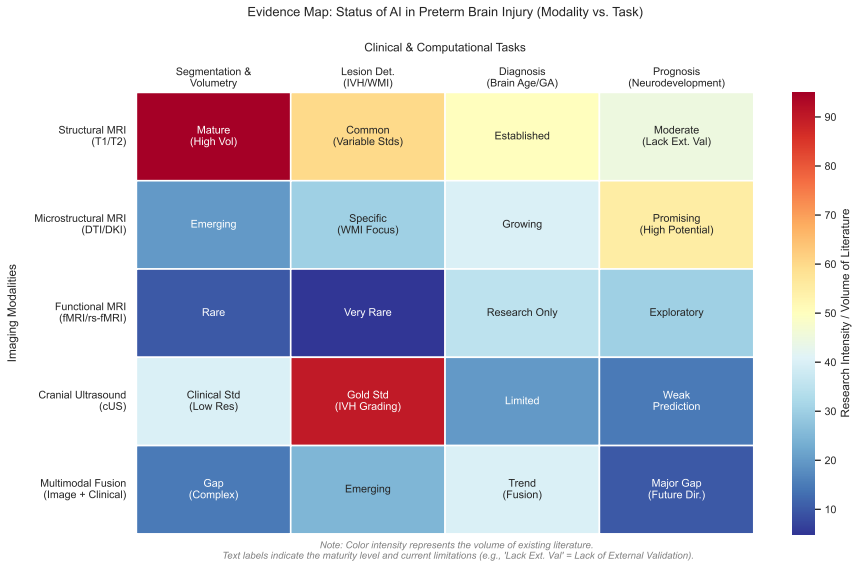

Figure 2. Evidence map by modality and task. Structural MRI segmentation has the strongest evidence, while multimodal fusion, functional imaging, prognosis, and external validation remain limited.

5. Barriers and minimum deployable pipeline

5.1. Barriers to trustworthy translation

Translation is constrained by mismatches between curated research data and routine neonatal care. Developmental appearance changes over weeks, while scanner protocols, reconstruction settings, ultrasound devices, and operator technique vary between sites. A model may therefore encounter a combination of anatomy and acquisition that was rare or absent during training. MRI availability is also related to physiological stability: infants who cannot be transported are not a random missing-data group. Treating their absent examination as an ordinary blank input can introduce systematic bias.

Reference labels create a second limitation. Tissue boundaries and diffuse abnormalities may have substantial reader variation, whereas rare but consequential lesions produce severe class imbalance. Consensus annotation can improve consistency but does not eliminate uncertainty about evolving injury. Dataset construction should record label source, reader expertise, adjudication, and the time between imaging and clinical events. Otherwise, an apparently precise model may reproduce an unstable definition.

DOI: 10.54254/2755-2721/2026.36134

Operational constraints are equally important. cUS triage may need results during acquisition, while MRI segmentation can run before scheduled reporting. Both require traceable preprocessing and model versions, a defined response when input quality is inadequate, and clear responsibility for accepting or overriding an output. Interfaces that display a confident result without its limitations can promote automation bias. Retrospective accuracy is therefore only one component of readiness; turnaround, failure behavior, auditability, and compatibility with local responsibility determine whether performance can transfer safely to a new neonatal unit.

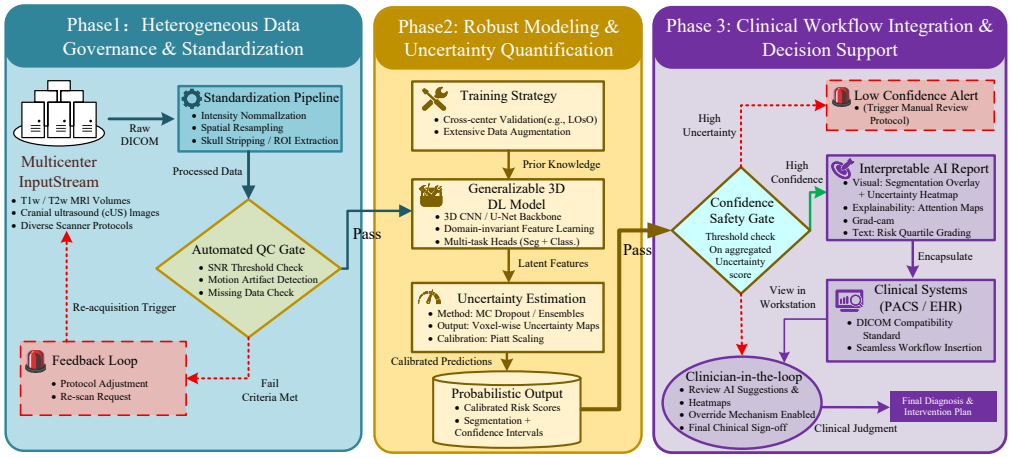

Figure 3. Minimum deployable pipeline for neonatal imaging AI. Quality control precedes inference; calibration and uncertainty accompany each prediction; and a safety gate routes low-quality or low-confidence cases to human review.

5.2. Minimum deployable pipeline

As shown in Figure 3, the minimum deployable pipeline has three linked phases. First, input governance verifies modality, sequence, spatial resolution, required views, motion, signal quality, and missing data before inference. Examinations outside prespecified limits should be rejected or sent for reacquisition rather than forced through the model [22]. The same checks must be applied during development and deployment so that quality rules do not change silently.

Second, robust inference combines infant-level testing with calibrated probabilities or uncertainty estimates. Validation on a held-out center should use a locked model and thresholds; uncertainty should define when the system abstains rather than serve only as an additional visualization [21, 23]. Third, the clinical interface presents the relevant measurement, lesion location, confidence, and comparison with prior imaging. Low-quality or low-confidence cases are routed to human review, and every output, override, and software version is logged. Monitoring after release should track input drift, abstention frequency, turnaround, subgroup performance, and clinically important errors. In this pipeline, abstention and a defined failure pathway are part of the clinical output, not signs that the model is incomplete.

Presentation should be task-specific. A cUS triage tool may display the acquired view, suspected abnormality, ventricular measurement, and change from the previous examination while the operator is still at the bedside. An MRI tool can return tissue or lesion overlays, regional volumes, quality flags, and developmental reference values before formal reporting. These outputs should enter the existing picture archiving or electronic record workflow rather than require a separate research interface. Interpretability should emphasize anatomically localized findings and quantitative comparison, not a

DOI: 10.54254/2755-2721/2026.36134

generic heatmap. Corrections made during review can support audit and later quality improvement, but they should not update a deployed model without controlled versioning and revalidation.

6. Discussion and future opportunities

For imaging and clinical tasks, future cohorts should be organized around decisions rather than around whichever sequence is available. A longitudinal design could link early serial cUS, term-equivalent structural and diffusion MRI, treatment exposures, and standardized developmental follow-up. Imaging time should be represented relative to gestational and postmenstrual age, and the reason an examination was omitted should be retained. Such cohorts would support separate questions: early detection of IVH or ventricular expansion, quantification of diffuse injury near term, and prediction of later impairment. Multimodal models should be required to outperform clinical-only and single-modality baselines and should demonstrate whether earlier information changes the timing or confidence of a decision.

For methods and evidence, limited annotation should motivate better use of unlabeled data rather than increasingly complex models trained on the same small cohorts. Self-supervised temporal or cross-modal pretraining could learn neonatal representations before task-specific fine-tuning, while semi-supervised methods could use examinations without complete lesion labels. Federated analysis may widen institutional representation when data cannot be centralized, but harmonized definitions and independent testing remain necessary. Models should be assessed on held-out centers with preregistered endpoints, subgroup calibration, and sensitivity analyses for missing modalities. Generative augmentation requires special caution: synthetic examples should be evaluated for anatomical fidelity, lesion preservation, and the possibility that generated patterns make classification artificially easy. Architecture comparisons should control preprocessing, splits, and baselines so that improvement can be attributed to the model rather than to study design.

For clinical translation, evaluation should proceed in stages. After retrospective multicenter testing, silent deployment can measure data availability, failure frequency, and turnaround without influencing care. Reader studies can then determine whether AI improves sensitivity, agreement, or reporting efficiency and whether benefit differs by experience. Prospective impact studies should assess management changes, unnecessary follow-up, missed injuries, workload, and automation-related errors, not only model accuracy. Protocols should prespecify who reviews abstained cases, how updates are controlled, and when performance deterioration triggers suspension. Patient and family communication also requires careful framing because prognostic probabilities are not deterministic outcomes. The most valuable future system will therefore be one that improves a defined neonatal workflow while preserving expert judgment and transparent responsibility.

7. Conclusion

This focused review identifies structural MRI segmentation as the application closest to routine use, while cUS triage, subtle-lesion detection, and multimodal prognosis remain less mature. The most practical starting point is a decision-specific workflow that combines modality-appropriate quality control with infant-level external validation and transparent human review. Future studies should model developmental time, preserve clinically informative missingness, and compare AI against both clinical and single-modality baselines. Translation should then progress through silent deployment, reader studies, and prospective multicenter impact evaluation using calibration, reporting time, workload, management change, and safety errors as endpoints. In this way, AI can support rather than replace neonatal expertise. The immediate research priority is therefore not another isolated high-accuracy model, but an auditable, uncertainty-aware pipeline that connects serial cUS and term-equivalent

DOI: 10.54254/2755-2721/2026.36134

MRI to reproducible decisions and long-term neurodevelopmental outcomes. This staged approach would provide stronger evidence that technical performance produces safe, equitable, and clinically meaningful benefit for preterm infants.

References

[1] Joseph J. Volpe. Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances. *The Lancet Neurology*, 8(1):110–124, 2009.

[2] Serafina Perrone, Federica Grassi, Chiara Caporilli, Giovanni Boscarino, Giulia Carbone, Chiara Petrolini, et al. Brain damage in preterm and full-term neonates: Serum biomarkers for the early diagnosis and intervention. *Antioxidants*, 12(2):309, 2023.

[3] Minhui Ouyang, Jessica Dubois, Qinlin Yu, Pratik Mukherjee, and Hao Huang. Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. *NeuroImage*, 185:836–850, 2019.

[4] Rebecca C. Knickmeyer, Sylvain Gouttard, Chaeryon Kang, Dianne Evans, Kathy Wilber, J. Keith Smith, et al. A structural MRI study of human brain development from birth to 2 years. *The Journal of Neuroscience*, 28(47): 12176–12182, 2008.

[5] Lianne J. Woodward, Peter J. Anderson, Nicola C. Austin, Kelly Howard, and Terrie E. Inder. Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *The New England Journal of Medicine*, 355(7):685–694, 2006.

[6] Lukun Tang, Qi Li, Feifan Xiao, Yanyan Gao, Peng Zhang, Guoqiang Cheng, et al. Neurosonography: Shaping the future of neuroprotection strategies in extremely preterm infants. *Heliyon*, 10(11):e31742, 2024.

[7] Thais Agut, Ana Alarcon, Fernando Cabanas, Marco Bartocci, Miriam Martinez-Biarge, and Sandra Horsch. Preterm white matter injury: Ultrasound diagnosis and classification. *Pediatric Research*, 87(S1):37–49, 2020.

[8] Jessica Dubois, Marianne Alison, Serena J. Counsell, Lucie Hertz-Pannier, Petra S. Hüppi, and Manon J.N.L. Benders. MRI of the neonatal brain: A review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging*, 53(5):1318–1343, 2021.

[9] Revital Nossin-Manor, Dallas Card, Drew Morris, Salma Noormohamed, Manohar M. Shroff, Hilary E. Whyte, et al. Quantitative MRI in the very preterm brain: Assessing tissue organization and myelination using magnetization transfer, diffusion tensor and T1 imaging. *NeuroImage*, 64:505–516, 2013.

[10] David Raffelt, J.-Donald Tournier, Stephen Rose, Gerard R. Ridgway, Robert Henderson, Stuart Crozier, et al. Apparent fibre density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage*, 59(4):3976–3994, 2012.

[11] Zhiyuan Li, Hailong Li, Anca L. Ralescu, Jonathan R. Dillman, Mekibib Altaye, Kim M. Cecil, et al. Joint self-supervised and supervised contrastive learning for multimodal MRI data: Towards predicting abnormal neurodevelopment. *Artificial Intelligence in Medicine*, 157:102993, 2024.

[12] M. Jorge Cardoso, Andrew Melbourne, Giles S. Kendall, Marc Modat, Nicola J. Robertson, Neil Marlow, et al. AdaPT: An adaptive preterm segmentation algorithm for neonatal brain MRI. *NeuroImage*, 65:97–108, 2013.

[13] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015.

[14] Hao Chen, Qi Dou, Lequan Yu, Jing Qin, and Pheng-Ann Heng. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170:446–455, 2018.

[15] Hailong Li, Jinghua Wang, Ming Chen, Venkata Sita Priyanka Illapani, Nehal A Parikh, and Lili He. Automatic segmentation of diffuse white matter abnormality on T2-weighted brain MR images using deep learning in very preterm infants. *Radiology: Artificial Intelligence*, 3(3):e200166, 2021.

[16] Yu Li, Xin Zhang, Jingxin Nie, Guowei Zhang, Ruiyan Fang, Xiangmin Xu, et al. Brain connectivity based graph convolutional networks and its application to infant age prediction. *IEEE Transactions on Medical Imaging*, 41(10): 2764–2776, 2022.

[17] Zehua Ren, Yongheng Sun, Miaomiao Wang, Yuying Feng, Xianjun Li, Chao Jin, et al. Punctate white matter lesion segmentation in preterm infants powered by counterfactually generative learning. In *Medical Image Computing and Computer Assisted Intervention*, volume 14224 of *Lecture Notes in Computer Science*, 2023.

[18] Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, et al. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

[19] Hailong Li, Junqi Wang, Zhiyuan Li, Kim M. Cecil, Mekibib Altaye, Jonathan R. Dillman, et al. Supervised contrastive learning enhances graph convolutional networks for predicting neurodevelopmental deficits in very

DOI: 10.54254/2755-2721/2026.36134

preterm infants using brain structural connectome. *NeuroImage*, 291:120579, 2024.

[20] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. SegMamba: Long-range sequential modeling mamba for 3D medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention*, volume 15001, pages 458–468, 2024.

[21] Tingting Huang, Jie Zheng, Heng Liu, Haoxiang Jiang, Chao Jin, Xianjun Li, et al. Development and validation of a conventional MRI-based model to predict cerebral palsy in infants (aged 6–24 months) with periventricular white matter injury: A multicentre, retrospective cohort study. *eClinicalMedicine*, 86:103364, 2025.

[22] Yijing Fang, Shirui Wang, Yihan Zhang, Chengxiu Yuan, Li Song, et al. A review of deep learning-based segmentation and registration of infant brain MRI. *Biomed. Signal Process. Control*, 112:108676, 2026.

[23] Jose Dolz, Christian Desrosiers, Li Wang, Jing Yuan, Dinggang Shen, and Ismail Ben Ayed. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, 79:101660, 2020.