BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE): Part I. MRI and Manual Lesion Annotation

Rina Bao[1,2,#], Ya'nan Song[1], Sara V. Bates[3], Rebecca J. Weiss[1], Anna N. Foster[1,3], Camilo Jaimes Cobos[1], Susan Sotardi[1], Yue Zhang[1], Randy L. Gollub[4,5,*], P. Ellen Grant[1,2,*], and Yangming Ou[1,2,*,#]

[1]Boston Children's Hospital, Boston, MA, USA
[2]Harvard Medical School, Boston, MA, USA
[3]Massachusetts General Hospital, Boston, MA, USA
[4]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA
[5]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[*]these authors contributed equally (RLG, PEG, YO)
[#]corresponding author(s): (rina.bao@childrens.harvard.edu, yangming.ou@childrens.harvard.edu)

ABSTRACT

Hypoxic ischemic encephalopathy (HIE) is a brain injury that occurs in $1 \sim 5/1000$ term neonates. Accurate identification and segmentation of HIE-related lesions in neonatal brain magnetic resonance images (MRIs) is the first step toward predicting prognosis, identifying high-risk patients, and evaluating treatment effects. It will lead to a more accurate estimation of prognosis, a better understanding of neurological symptoms, and a timely prediction of response to therapy. We release the first public dataset containing neonatal brain diffusion MRI and expert annotation of lesions from 133 patients diagnosed with HIE. HIE-related lesions in brain MRI are often diffuse (i.e., multi-focal), and small (over half the patients in our data having lesions occupying <1% of brain volume). Segmentation for HIE MRI data is remarkably different from, and arguably more challenging than, other segmentation tasks such as brain tumors with focal and relatively large lesions. We hope that this dataset can help fuel the development of MRI lesion segmentation methods for HIE and small diffuse lesions in general.

Background & Summary (700 words maximum)

Accurate identification of brain lesion injuries in neonatal brain magnetic resonance images (MRI) [1, 2, 3] is crucial to improve clinical care of neonates with hypoxic ischemic encephalopathy (HIE), a brain disease that occurs in around $1 \sim 5/1000$ term-born infants at birth [4, 5]. HIE affects around 200,000 term-born neonates every year worldwide [4, 5], costing about $2 billion/year in the US alone, let alone family burdens. Therapeutic hypothermia, the current clinical treatment of HIE, can reduce mortality and morbidity. Nevertheless, around 60% of patients still die or develop neurocognitive deficits by 2 years of age. MRI is used in over 50% of all the >100 ongoing HIE-related clinical trials worldwide [6], for evaluating treatment effects [7, 8, 9], and helping discover clinical [10, 11, 12], biochemical [10, 13, 14, 15], and serum [16, 17, 18] biomarkers. Accurate identification of brain lesions in neonatal brain MRIs [1, 2, 3] is needed for disease prognosis, a better understanding of the neural basis of disease progression, and more timely evaluations of therapeutic effects.

HIE lesions are often diffuse (i.e., multi-focal), and small; hence, algorithms that have shown great promise in segmenting big and focal lesions, such as brain tumors and acute strokes, often encounter challenges when directly applied to MRIs of HIE patients. Indeed, many (over half) patients had lesions occupying <1% of brain volume, as shown in Figure 1. As a result, the segmentation accuracy measured by the Dice overlap with U-Net [19] and other state-of-the-art machine/deep learning algorithms on HIE remains at around 0.5 [20], whereas Dice is over 0.8 when segmenting brain tumors [21, 22].

A major hurdle in developing algorithms for small diffuse lesions, such as HIE lesions, is the lack of public data. Public data with expert annotations of lesions have fueled the advancement of machine learning algorithms to segment brain tumors [23], stroke lesions [24], multiple sclerosis lesions [25, 26], and numerous other diseases in the brain or other organs [27, 28]. However, to date, there is no public MRI data with expert annotations available for HIE lesions.

We present BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE), an open-source, comprehensive, and representative MRI dataset for HIE. This paper introduces the first part of the BONBID-HIE data. This release contains raw and derived diffusion parameter maps, as well as manually-annotated lesion masks, for 133 HIE patients.

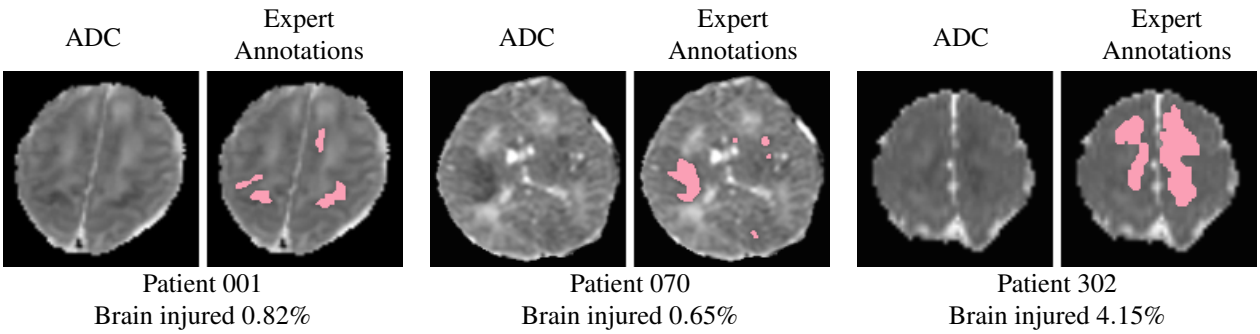


Figure 1. Lesions associated to hypoxic ischemic encephalopathy (HIE) are typically diffuse (i.e., multi-focal) and small. Here we show two representative images for 3 HIE patients. For each patient, in the left panel: apparent diffusion coefficient (ADC) maps that are clinically used to identify HIE lesions; in the right panel: manually-annotated lesions (shown in pink) overlaid on the ADC map. We listed the percentage of the whole brain volume being injured at the bottom (i.e., lesion volume divided by the whole brain volume).

Our data was from Massachusetts General Hospital. It includes MRIs from different scanners (Siemens 3T and GE 1.5T), different MRI protocols, and from patients of different races/ethnicities and ages (0-14 days postnatal age). Part I of our data release (this paper) focuses on lesion detection, while Part II (a follow-up paper) will focus on clinical, treatment, and neurologic outcome data for further developing prognostic biomarkers.

Methods

This work was approved by the Institutional Review Boards (IRBs) at Massachusetts General Hospital (MGH) and Boston Children's Hospital (BCH). Figure 2 illustrates the overall data archiving process.

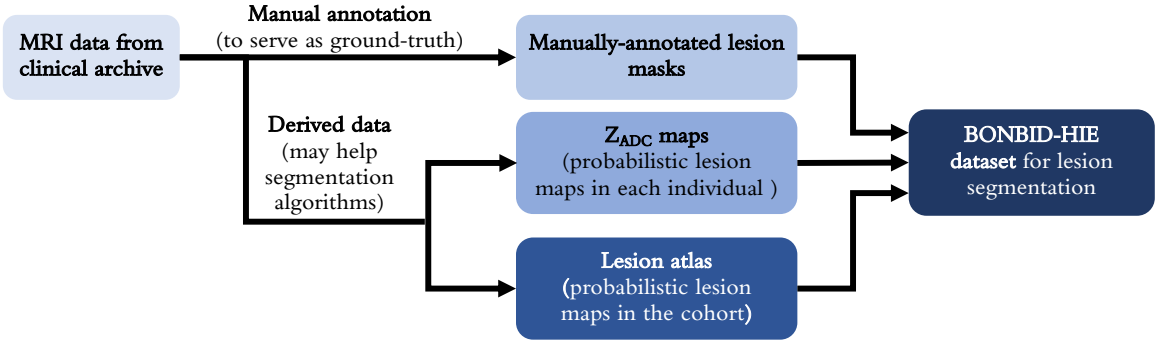


Figure 2. A schematic diagram showing the steps performed on the BONBID-HIE data for release.

Retrospective Data Collection

Data was retrospectively collected from MGH. Inclusion criteria were: (1) term-born (at physician discretion) (2) clinical diagnosis of HIE; (3) initially treated at MGH between 2001 and 2018; (4) no comorbidities such as hydrocephalus or congenital syndromes; and (5) high-quality MRI acquired in Day 0-14 after birth (visually checked by RW, AF, YO). Exclusion criteria were: (1) excessive motion artifacts or missing images; (2) secondary HIE diagnosis to a primary perinatal stroke.

Clinical characteristics and demographic information were retrospectively gathered from the electronic health records (EHRs). The clinical variables included maternal information during pregnancy and delivery, as well as infant information. More detail can be found in the "Data Records" Section.

MRI data was downloaded from MGH Radiology Department clinical archives using the mi2b2 search engine [29]. MRIs were acquired on either a GE 1.5T Signa scanner (N=52, scanned during 2001-2012), or, a Siemens 3T TrioTim or PrismaFit scanner (N=81, scanned during 2012-2018). Diffusion tensor sequences on all scanners had the protocol as follows: Time of Repetition (TR)=7500 – 9500ms, Time of Echo (TE)=80 – 115ms, and b=1000 s/mm^2. The GE scanner had resolution $1.5 \times 1.5 \times (2.0 - 4.0) mm^3$ and (6 - 60) diffusion directions, while the Siemens scanner had a resolution $2 \times 2 \times 2 mm^3$ and (25

- 60) diffusion directions. Apparent diffusion coefficient (ADC) maps were directly generated by the scanners (with Syngo software for Siemens scanners [30], and with the Advantage Windows Workstation for GE scanners [31, 32]).

MRI Pre-processing

Besides the raw NIfTI image as converted from the DICOM files, we also generated several processed images. The pre-processing steps included: N4 bias correction [33], field of view normalization [34], multi-atlas skull stripping for the ADC maps [35], and deformable registration of each patient's ADC map to a normative 0-14 day neonatal brain ADC atlas [36], by the Deformable Registration via Attribute Matching and Mutual-Saliency weighting (DRAMMS) software [37], which has been extended-validated for lifespan ages in various MRI sequences [38]. This normative ADC atlas was constructed from ADC maps of 13 healthy individuals acquired 0-14 days after birth (Figure 3a) with our extensively-validated MRI analysis pipeline [39, 38, 35, 34]. All software packages used in this pre-processing pipeline are publicly available and have been validated in processing both research and clinical MRI scans across ages [40, 41, 42, 43].

Expert Annotation of Lesions

HIE lesions were manually annotated as a binary mask on the 3D ADC maps in the patient's raw image space, using the MRICroN software. ADC maps are clinically used as the standard images to identify HIE abnormalities [44, 45]. The annotations were done by a trained physician (YS; >3 years of experience) according to the neuroradiology reports that were generated as part of the clinical flow. The annotations started from the axial slice and were subsequently modified in the coronal and sagittal planes for the 3D integrity of lesion regions. Uncertainties occurred in 27 patients and were resolved by the consensus of three pediatric neuroradiologists (CJ, SS, and PEG; >5, >5, and >20 years of experience).

Generation of Z_{ADC} Maps for Each Patient

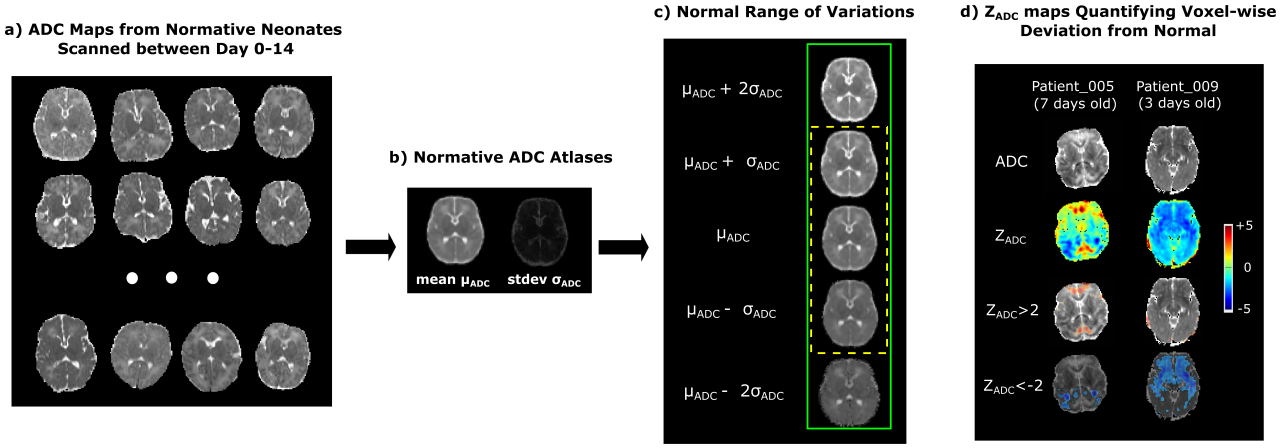


Figure 3. The generation and concept of Z_{ADC} maps. (a) Examples of ADC maps from normative subjects, which were warped into the same space using unbiased group-wise DRAMMS registration to generate (b) the mean and standard deviation ADC atlases. (c) One or two standard deviations above and below the mean ADC atlas define the normal ranges of voxel-wise ADC variations. (d) Our novel Z_{ADC} map quantifies voxel-wise deviations from the mean ADC map in (b). The cool/warm colors in (d) represent voxels with ADC values lower/higher than the mean ADC at the same anatomic location, according to the scale bar on the right.

Neuroradiologists identify acute brain injury from HIE as regions with low ADC values. Low ADC values represent a reduced water diffusion or blood supply [6]. However, a dilemma is, what ADC value is considered abnormally low versus just low within the normal variation? The normal variations of ADC values differ across brain regions [36, 46], making this question difficult even for experienced neuroradiologists. For example, a voxel with an ADC value of 800 ($\times 10^{-6} mm^2/s$) may be considered normal at one brain region, whereas another voxel with an ADC value of 900 ($\times 10^{-6} mm^2/s$) may be considered lesioned at another brain region, if the normal ranges of ADC variations in the two brain regions are 700-900 and 950-1100 (same unit), respectively.

To address this dilemma, we have developed Z_{ADC} maps to normalize and make ADC values comparable across brain voxel locations [40]. First, a normative ADC atlas was generated from scans of healthy neonates (Figure 3a). This atlas quantifies the mean ADC values and standard deviation at every voxel [47] (Figure 3b), and hence the normal range of variations at each

voxel (Figure 3c). Then, we converted each patient's ADC map (first row, Figure 3d) into a Z_{ADC} map (second row, Figure 3d). The Z_{ADC} maps compared the patient's ADC value at each voxel to the normal variations at the corresponding voxel in the atlas. In a nutshell, Z_{ADC} maps quantify how many standard deviations away a patient's ADC value at a voxel is from the normal mean at the same anatomical location.

Specifically, a deformation (D) was computed, which mapped every voxel x in the patient's ADC map to its anatomically-corresponding location $D(x)$ in the atlas space. The normal range of ADC variation per voxel was defined by the mean and standard deviation denoted for that voxel across all healthy neonates. Finally, the patient's ADC value I_x at voxel x was converted to a Z value: $Z_x = (I_x - \mu D(x))/\sigma D(x)$. We calculated the Z_{ADC} map, which resides in the patient's raw ADC image space, for each patient. This offers an option for developing anatomy-aware lesion segmentation algorithms [48].

Construction of Statistical Lesion Atlases for the Cohort

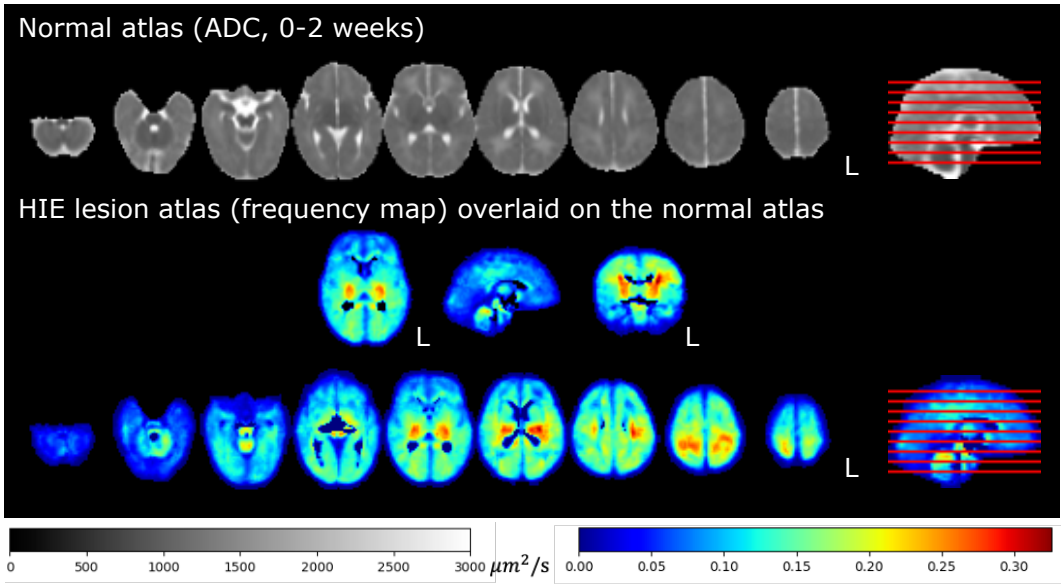


Figure 4. Statistical lesion atlas quantifying the voxel-wise lesion frequency in our cohort of N=133 patients in the normal 0-14 days ADC atlas space.

The same deformation field that was computed by the non-rigid registration from the patient's skull-stripped ADC map to the normal ADC atlas was used to transform the binary brain lesion maps of each patient into the normal neonatal ADC atlas space [37]. The transformed binary lesion masks were then summed and divided by the total number of patients at each voxel. This led to a statistical lesion atlas that quantifies voxel-wise frequency, or probability, of HIE lesions in our cohort, as illustrated in Figure 4.

Data Records

Dataset Characteristics

Table 1A lists the demographics and clinical characteristics of mothers and neonates. Maternal information includes demographics (age at delivery, race), birth mode (C-section or vaginal), and complications during pregnancy and delivery. Neonatal information includes demographics (age at MRI scan, gestational age at birth, birth weight, head circumference, sex), birth conditions (1/5/10-minute APGAR scores, lowest pH value in umbilical cord), treatment (hypothermia or not), and complications in the neonatal intensive care unit (NICU), including seizure (yes/no), length of stay (in days), the use of endotracheal tube (ETT, yes/no), and the administration of total parenteral nutrition (TPN, yes/no). In each row, we also listed the number of patients who had such information available.

Table 1B quantifies the distribution of the absolute lesion volumes (in mm^3) and relative lesion volume (percentage of the brain being injured). Here, the relative lesion volume was calculated by the volume of the expert-annotated lesion regions divided by the algorithm-extracted whole brain volumes [34, 35]. The median lesion volume accounted for 0.63% of the whole brain volume. This confirms that over half of the patients had less than 1% of the brain being injured. Table 1C further calculates the distribution of the relative lesion volumes (by the percentage of the brain being lesioned). The absolute and

relative volumes were both computed in the patient's raw ADC image space. The minimum lesion was 0 mm^3, which is a common issue in HIE – mild HIE cases may not show explicit lesions in neonatal MRIs [49, 50, 51].

Figure 5 shows the ADC map, Z_{ADC} map, and expert annotations of example patients with different HIE lesion percentages. Two patients are shown in each of the four groups: those with lesions occupying <1% (upper left panel), 1-5% (upper right panel), 5-50% (lower left panel), and 50-100% (lower right panel) of the whole brain volume. Overall, around 1 in 2 (55.64%) patients had HIE lesions occupying less than 1% of their brain volume, and 3 in 4 patients (75.19%) patients had lesions occupying less than 5% of their brain. This confirms that HIE lesions detectable in the diffusion MRI in our cohort are often small.

Data structure and file formats

All medical imaging files were exported from the Picture Archiving and Communication System (PACS) and converted into the NIfTI format. Segmentation masks created by expert annotations were also saved in NIfTI format. Corresponding scanner metadata from the Digital Imaging and Communications in Medicine (DICOM) header in the .json file format is provided with the datasets. All data in the BONBID-HIE dataset was separated into a training dataset (n=89) and a test dataset (n=44). Both the training and test sets contain data from both scanners (GE 1.5T Signa and Siemens 3T Trio). The split between the training and test data set has been performed (RB, YO) so that both sets include a similar variance of HIE lesion patterns as shown in Table 1C.

Table 1. Cohort characteristics (N=133)

A. Demographics and Clinical Characteristics		
Maternal Information		
Maternal age at delivery (years)	29.5 ± 6.7	N=133
Race	White (43), Black or African American (7), Hispanic or Latino (15), Multi Race (5), Unknown (57), Other (6)	N=133
Delivery	C-section (78), Vaginal (55)	N=133
Antepartum hemorrhage	Yes (29), No (104)	N=133
Thyroid dysfunction	Yes (5), No (128)	N=133
Pre-eclampsia	Yes (9), No (124)	N=133
Fetal decels	Yes (72), No (61)	N=133
Shoulder dystocia	Yes (8), No (125)	N=133
Chorioamnionitis	Yes (20), No (108)	N=133
Emergency c-section	Yes (69), No (58)	N=133
Neonatal Information		
Age at scan (days)	3.9 ± 2.7	N=133
Gestational age at birth (weeks)	39.1 ± 1.9	N=133
Birth weight (g)	3321.9 ± 615.9	N=133
Infant head circumference (cm)	34.2 ± 1.4	N=85
Sex	Male (74), Female (59)	N=133
1-minute APGAR scores	1.9 ± 1.7	N=133
5-minute APGAR scores	4.2 ± 2.3	N=132
10-minute APGAR scores	5.3 ± 2.1	N=118
Lowest pH value in umbilical cord	7.00 ± 0.20	N=129
Therapeutic hypothermia before MRI?	Yes (86), No (47)	N=133
Endotracheal tube (ETT) in NICU	Yes (78), No (47)	N=125
Total parenteral nutrition (TPN) in NICU	Yes (111), No (21)	N=133
Seizures NICU	Yes (64), No (69)	N=133
Length of stay in NICU (days)	12.10 ± 9.92	N=133
B. Lesion Characteristics (N=133)		
Whole-brain lesion volume – minimum	0 mm3 (0% of the brain injured)	
Whole-brain lesion volume – 25th percentile	441.96 mm3 (0.10% of the brain injured)	
Whole-brain lesion volume – median	2765.63 mm3 (0.63% of the brain injured)	
Whole-brain lesion volume – 75th percentile	24264.27 mm3 (4.86% of the brain injured)	
Whole-brain lesion volume – maximum	412120.00 mm3 (82.59% of the brain injured)	
C. Number of Patients by Percentage of Lesions in the Brain (N=133)		
$[0\%, 1\%)$ of the brain being injured	55.64% (N=74)	
$[1, 5\%)$ of the brain was injured	19.55% (N=26)	
$[5, 10\%)$ of the brain was injured	5.26% (N=7)	
$[10, 20\%)$ of the brain was injured	4.51% (N=6)	
$[20, 50\%)$ of the brain was injured	8.27% (N=11)	
$[50, 100]\%$ of the brain was injured	6.77% (N=9)	
D. Scanner (N=133)		
GE 1.5T	39% (N=52)	
Siemens 3T	61% (N=81)	

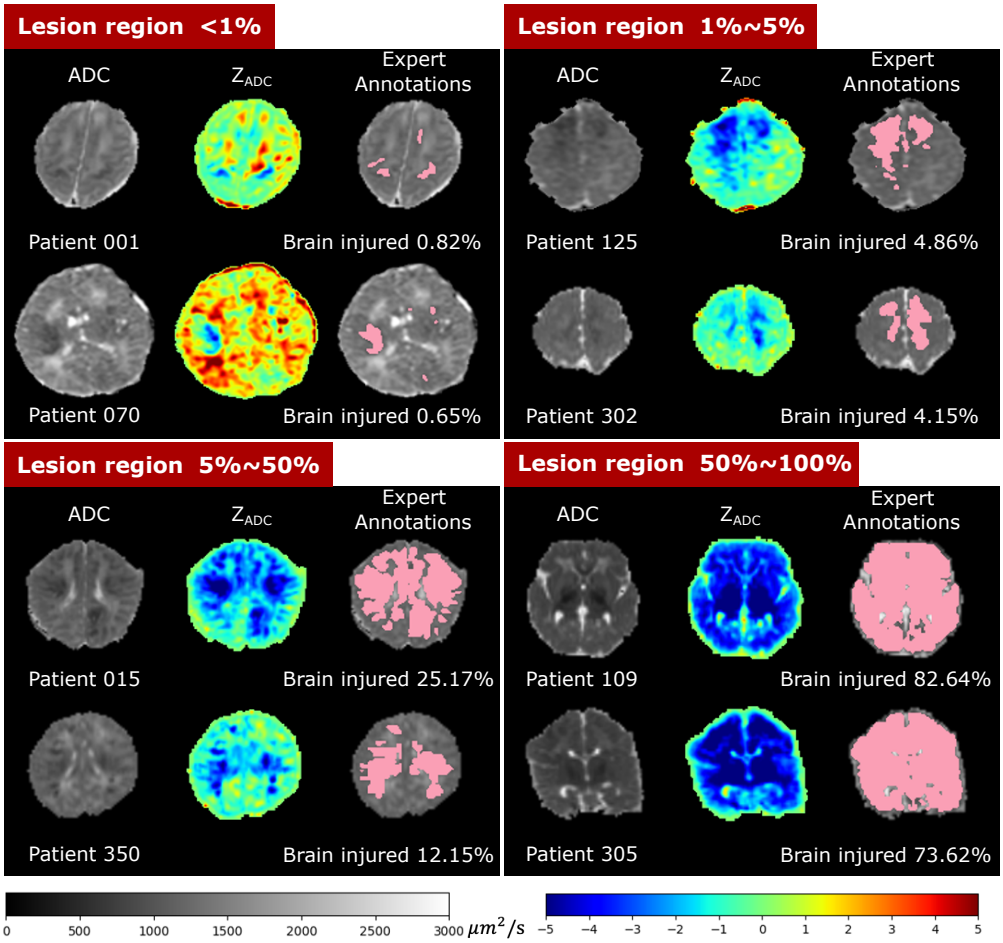


Figure 5. Visualization of patients with different lesion percentages. In every patient, the left image is the ADC map (skull stripped) with range of ADC values designated by the gray scale bar, the middle is the computed Z_{ADC} map with range of Z-scores designated by the rainbow scale bar, and the right image is the expert-annotated lesion regions (pink) overlaid on the ADC map. Percentages of injury were calculated by the volume of the expert-annotated lesion regions divided by algorithm-extracted whole brain volumes.

The data is organized in the format shown in Figure 6. BONBID-HIE provides, per patient: (i) 0ADC: raw defaced ADC maps; (ii) 1ADC_ss: skull stripped ADC map; (iii) 2Z_ADC: Z_{ADC} map; (iv) 3LABEL: expert lesion annotations; and (v) clinical data: clinical variables as written in Table 1A. There is also (vi) Atlases: a folder for the normal and lesion atlases; (vii) a readme.txt file: a text file to provide information on this data organization; and (viii) the license file of the BONBID-HIE dataset.

Data repository and storage

All training data has been made publicly available under the CC BY NC ND license (https://creativecommons.org/licenses/by-nc-nd/2.0/, allowing academic use with credit, prohibiting commercial use without owner's permission, and disallowing derivation or adaption of data). This dataset is also used for the HIE lesion segmentation challenge in Medical Image Computing and Computer Assisted Intervention (MICCAI) 2023 annual conference in October 2023. Further information about the HIE challenges, including data storage and download, can be found under our homepage (https://hiechallenge.github.io/). As such, the expert lesion annotation of lesions in the testing sub-cohort will be released after the challenge.

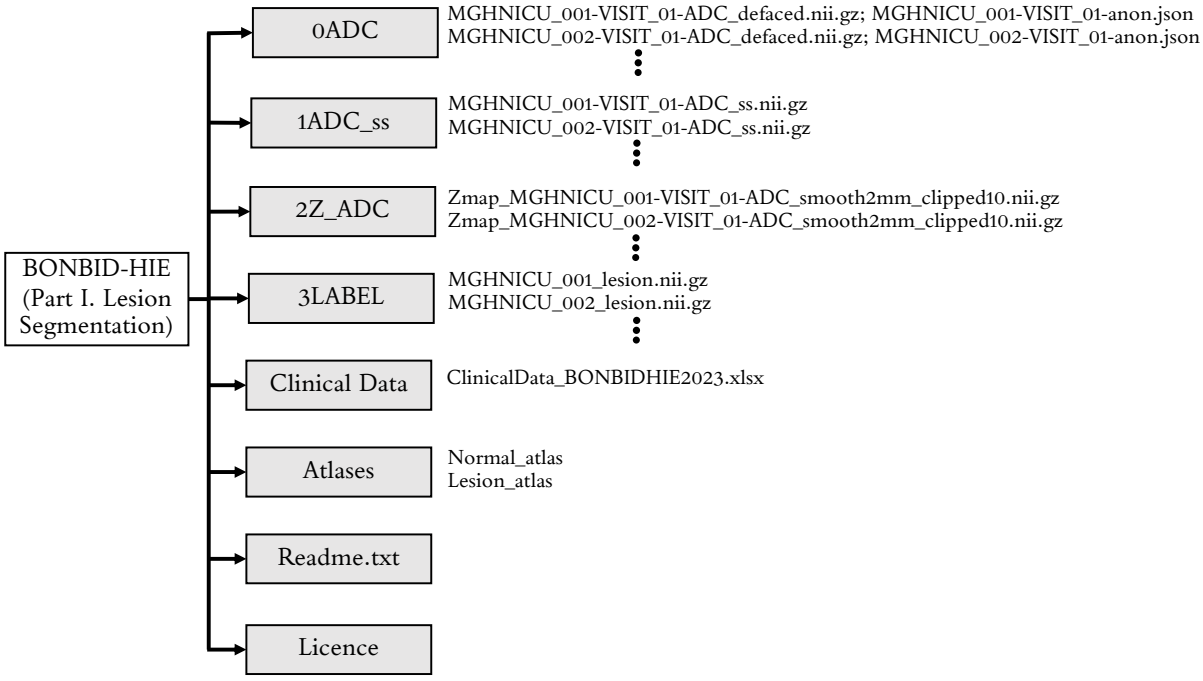


Figure 6. Folder structure of the BONBID-HIE dataset (Part I. Lesion Segmentation).

Technical Validation

Representativeness of patient cohort

Our data is representative of HIE cohorts in the developed countries. At least three characteristics of our data agree with documented clinical knowledge about HIE.

Lesion distribution in space agrees with clinical knowledge. Our statistical lesion atlas in Figure 4 shows that HIE lesions can occur anywhere in the brain. The regions most frequently injured here included the basal ganglia, internal capsules, thalamus, temporal lobes, cerebral white matter, brainstem, and cerebellum (red, orange, and yellow regions in Figure 4). This lesion atlas map coincides with clinical knowledge of brain regions often vulnerable to HIE injuries [2, 52, 6]. Indeed, HIE-related injuries in these regions have been key criteria in expert MRI scoring systems, which are used to assess the severity of HIE. Examples include the NICHD Neonatal Research Network (NRN) [2], the Barkovich [53], the Weeke-deVeries [3], and the Trivedi [54] scoring systems. In addition, lesions appeared in less than 35% of the patients at any given voxel, according to the color bar in this figure. This confirms the clinical knowledge that HIE lesions are diffuse, spatially distributed, and almost half to two-thirds of the HIE patients are mild to moderate, at least in patients in the USA [50, 51].

Lesion distribution in time agrees with clinical knowledge. Figure 7(a) shows the percentage of the whole brain volume being lesioned at different postnatal ages. The lesion percentage in the ADC maps came down to almost 0 in the 9 patients who underwent MRI scans after postnatal day 7. This agrees with the clinical knowledge that HIE-related lesions are more detectable in ADC during 0-7 postnatal days or in T1/T2-weighted images than in relatively later scans (after postnatal day 7) [55, 51, 2].

ADC evolvement with age agrees with clinical knowledge. Figure 7(b) shows the whole-brain average ADC values of all patients (each dot is a patient). In normal cohorts, ADC values drop rapidly in the early postnatal life (see Figures 4 and 5 in [36], and Figures 3 and 4 in [46]). However, the presence of HIE-related abnormalities disrupted this trend – HIE patients undergoing earlier MRIs (0-7 postnatal days) had decreased ADC values in a larger precentage of the brain (Figure 7(a)), so the ADC values in 0-7 postnatal days were at similar or even lower levels than ADC values in 7-14 postnatal days among HIE patients (Figure 7(b)). This has also been documented in HIE literature [56, 57].

Utility of ADC maps and Z_{ADC} maps

To demonstrate the utility of the computed Z_{ADC} maps, we compared the accuracies of using ADC or Z_{ADC} maps for lesion segmentation. We attempted simple thresholding of ADC and Z_{ADC} maps, at several threshold values, for segmenting HIE lesions. Although simple, thresholding-based segmentation accuracy is a strong indicator for segmentation accuracies in more

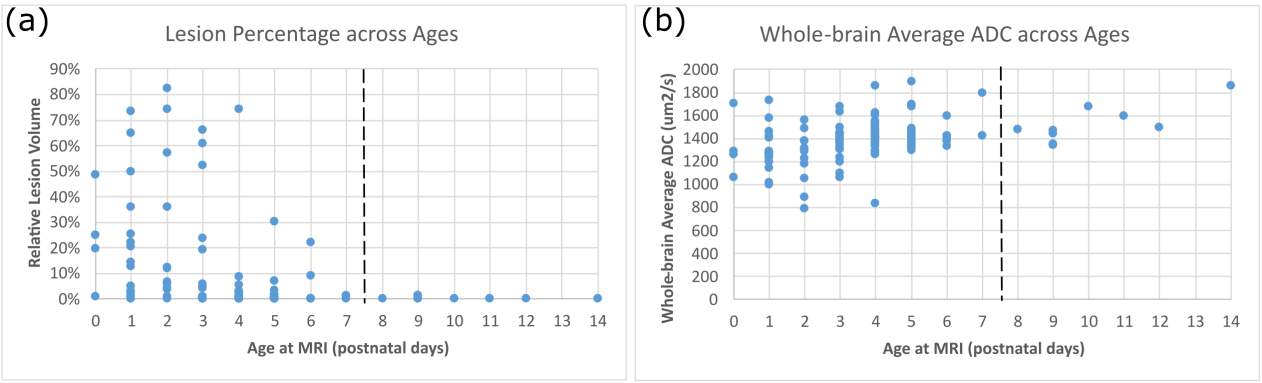


Figure 7. Representativeness of our cohort for (a) lesion distribution across ages; and (b) whole-brain ADC values across ages. In both panels, each dot denotes a patient in our cohort.

sophisticated machine/deep learning algorithms [58]. For ADC maps, we used different thresholds ranging from 800-1100 $\mu m^2/s$, as suggested in the literature [55, 59, 60, 61, 62]. For Z_{ADC} maps, we used thresholds -1.5, -2, and -2.5. Voxels in patient MRIs with ADC values 1.5 to 2.5 standard deviations below the average ADC values from healthy controls were considered abnormally low and hence, lesioned. The choice of thresholds around -2 in Z_{ADC} maps was also based on the normal distribution of ADC values at each voxel across subjects [36, 40].

We evaluated the accuracy of these maps compared to expert-annotated ground-truth masks using the Dice coefficient, sensitivity, and specificity. Results are shown in Figure 8. Here, the gray boxplots are the accuracy measurements when ADC maps were thresholded between 800 and 1100 $\mu m^2/s$ (50 $\mu m^2/s$ intervals). The blue boxplots are the accuracy measurements when the Z_{ADC} maps were thresholded at -1.5, -2, and -2.5. Figure 8 demonstrates: (i) both ADC and Z_{ADC} have value in helping segment the HIE-related lesions, since the specificity from simple naive thresholding-based segmentations was comparable to those from machine learning-based algorithms [20], although the Dice and sensitivity were lower; (ii) Z_{ADC} maps thresholded at -2, the most intuitive and straightforward threshold value, yielded the highest Dice (0.54 ± 0.28), followed by Z_{ADC} maps thresholded at -2.5 (Dice 0.39 ± 0.25); (iii) Z_{ADC} maps thresholded at -1.5 yielded the highest $\frac{sensitivity^2 + specificity^2}{2}$ (0.82 ± 0.02) compared with any ADC thresholds, followed by Z_{ADC} maps thresholded at -2 (0.76 ± 0.03); and (iv) overall, across all thresholds, Z_{ADC} maps showed a higher area under the curve (AUC: 0.936). This shows that Z_{ADC} maps – anatomy-normalized ADC images – carry the potential to improve lesion detection accuracy.

Code Availability

Data is available as part of the 1st BONBID-HIE MICCAI challenge (https://hiechallenge.github.io/), to be held in October 2023. For the training sub-cohort, all MRI (raw and derived) and expert lesions annotations are available. For the testing sub-cohort, only the MRI (raw and derived) data is available at the time this manuscript is drafted, and the expert lesion annotations will be made available after the challenge in October 2023.

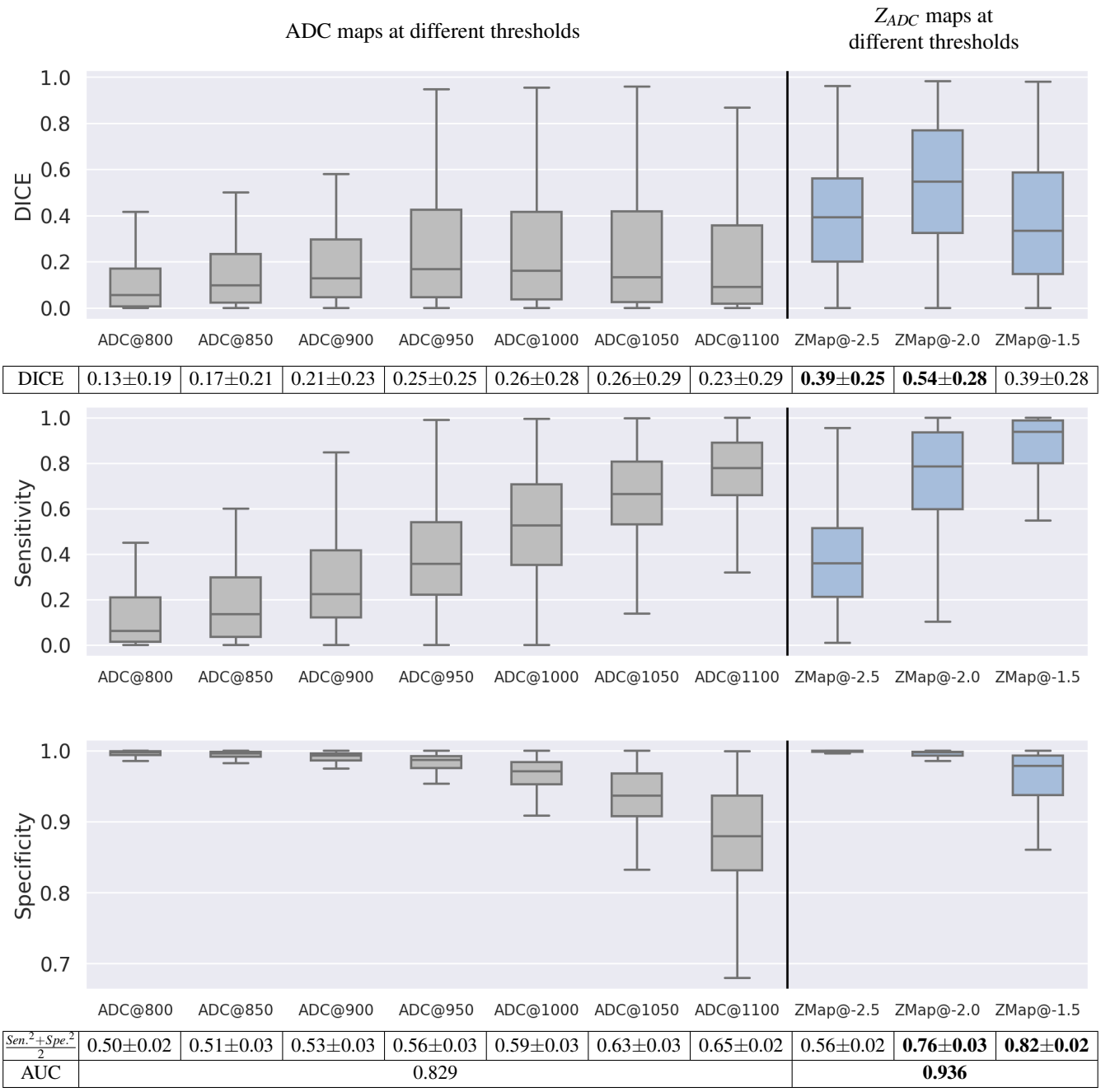
Codes to automatically calculate the evaluation metrics are available in BONBID-HIE 2023 MICCAI challenge (https://bonbid-hie2023.grand-challenge.org/). The repository in the challenge contains scripts to read the images, visualize them, and quantify the algorithm's performance with the same metrics used in the challenge to rank participants.

Acknowledgements

This work was supported, in part, by NIH R03HD104891, R21NS121735, and R61NS126792.

Competing interests

None.



DICE	0.13±0.19	0.17±0.21	0.21±0.23	0.25±0.25	0.26±0.28	0.26±0.29	0.23±0.29	**0.39±0.25**	**0.54±0.28**	0.39±0.28
$\frac{Sen.^2+Spe.^2}{2}$	0.50±0.02	0.51±0.03	0.53±0.03	0.56±0.03	0.59±0.03	0.63±0.03	0.65±0.02	0.56±0.02	**0.76±0.03**	**0.82±0.02**
AUC	0.829							**0.936**		

Figure 8. Accuracy of thresholding-based lesion segmentation on ADC and Z_{ADC} maps with different threshold values. Bold texts in the tables beneath the figure panels highlight the two scenarios with the highest Dice scores, the two scenarios with the highest and most balanced sensitivity and specificity metrics, and the parametric map (ADC or Z_{ADC}) with the highest area under the receiver-operating-characteristic curve (AUC).

References

[1] Mary Rutherford, Christina Malamateniou, Amy McGuinness, Joanna Allsop, Miriam Martinez Biarge, and Serena Counsell. "Magnetic resonance imaging in hypoxic-ischaemic encephalopathy". In: *Early human development* 86.6 (2010), pp. 351–360.

[2] Seetha Shankaran, Patrick D Barnes, Susan R Hintz, Abbott R Laptook, Kristin M Zaterka-Baxter, Scott A McDonald, Richard A Ehrenkranz, Michele C Walsh, Jon E Tyson, Edward F Donovan, et al. "Brain injury following trial of

hypothermia for neonatal hypoxic–ischaemic encephalopathy". In: *Archives of Disease in Childhood-Fetal and Neonatal Edition* 97.6 (2012), F398–F404.

[3] Lauren C Weeke, Floris Groenendaal, Kalyani Mudigonda, Mats Blennow, Maarten H Lequin, Linda C Meiners, Ingrid C van Haastert, Manon J Benders, Boubou Hallberg, and Linda S de Vries. "A novel magnetic resonance imaging score predicts neurodevelopmental outcome after perinatal asphyxia and therapeutic hypothermia". In: *The Journal of pediatrics* 192 (2018), pp. 33–40.

[4] Ernest M Graham, Kristy A Ruis, Adam L Hartman, Frances J Northington, and Harold E Fox. "A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy". In: *American journal of obstetrics and gynecology* 199.6 (2008), pp. 587–595.

[5] Anne CC Lee, Naoko Kozuki, Hannah Blencowe, Theo Vos, Adil Bahalim, Gary L Darmstadt, Susan Niermeyer, Matthew Ellis, Nicola J Robertson, Simon Cousens, et al. "Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990". In: *Pediatric research* 74.1 (2013), pp. 50–72.

[6] Rebecca J Weiss, Sara V Bates, Ya'nan Song, Yue Zhang, Emily M Herzberg, Yih-Chieh Chen, Maryann Gong, Isabel Chien, Lily Zhang, Shawn N Murphy, et al. "Mining multi-site clinical data to develop machine learning MRI biomarkers: application to neonatal hypoxic ischemic encephalopathy". In: *Journal of Translational Medicine* 17.1 (2019), pp. 1–16.

[7] Tamara I Herrera, Laura Edwards, William F Malcolm, P Brian Smith, Kimberley A Fisher, Carolyn Pizoli, Kathryn E Gustafson, Ricki F Goldstein, C Michael Cotten, Ronald N Goldberg, et al. "Outcomes of preterm infants treated with hypothermia for hypoxic-ischemic encephalopathy". In: *Early human development* 125 (2018), pp. 1–7.

[8] Terrie E Inder, Rod W Hunt, Colin J Morley, Lee Coleman, Michael Stewart, Lex W Doyle, and Susan E Jacobs. "Randomized trial of systemic hypothermia selectively protects the cortex on MRI in term hypoxic-ischemic encephalopathy". In: *The Journal of pediatrics* 145.6 (2004), pp. 835–837.

[9] Elizabeth E Rogers, Sonia L Bonifacio, Hannah C Glass, Sandra E Juul, Taeun Chang, Dennis E Mayock, David J Durand, Dongli Song, Anthony J Barkovich, Roberta A Ballard, et al. "Erythropoietin and hypothermia for hypoxic-ischemic encephalopathy". In: *Pediatric neurology* 51.5 (2014), pp. 657–662.

[10] Girija Natarajan, Athina Pappas, and Seetha Shankaran. "Outcomes in childhood following therapeutic hypothermia for neonatal hypoxic-ischemic encephalopathy (HIE)". In: *Seminars in perinatology*. Vol. 40. 8. Elsevier. 2016, pp. 549–555.

[11] Vijay Ramaswamy, Jennifer Horton, Ben Vandermeer, Nina Buscemi, Steven Miller, and Jerome Yager. "Systematic review of biomarkers of brain injury in term neonatal encephalopathy". In: *Pediatric neurology* 40.3 (2009), pp. 215–226.

[12] Henriette van Laerhoven, Timo R de Haan, Martin Offringa, Bart Post, and Johanna H van der Lee. "Prognostic tests in term neonates with hypoxic-ischemic encephalopathy: a systematic review". In: *Pediatrics* 131.1 (2013), pp. 88–98.

[13] Abbot R Laptook, Seetha Shankaran, Jon E Tyson, Breda Munoz, Edward F Bell, Ronald N Goldberg, Nehal A Parikh, Namasivayam Ambalavanan, Claudia Pedroza, Athina Pappas, et al. "Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial". In: *Jama* 318.16 (2017), pp. 1550–1560.

[14] Seetha Shankaran, Abbot R Laptook, Athina Pappas, Scott A McDonald, Abhik Das, Jon E Tyson, Brenda B Poindexter, Kurt Schibler, Edward F Bell, Roy J Heyne, et al. "Effect of depth and duration of cooling on death or disability at age 18 months among neonates with hypoxic-ischemic encephalopathy: a randomized clinical trial". In: *Jama* 318.1 (2017), pp. 57–67.

[15] Keliana O'Mara and Michael D Weiss. "Dexmedetomidine for sedation of neonates with HIE undergoing therapeutic hypothermia: a single-center experience". In: *American Journal of Perinatology Reports* 8.03 (2018), e168–e173.

[16] An N Massaro, Andreas Jeromin, Nadja Kadom, Gilbert Vezina, Ronald L Hayes, Kevin KW Wang, Jackson Streeter, and Michael V Johnston. "Serum biomarkers of MRI brain injury in neonatal hypoxic ischemic encephalopathy treated with whole-body hypothermia: a pilot study". In: *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 14.3 (2013), p. 310.

[17] An N Massaro, Yvonne W Wu, Theo K Bammler, Bryan Comstock, Amit Mathur, Robert C McKinstry, Taeun Chang, Dennis E Mayock, Sarah B Mulkey, Krisa Van Meurs, et al. "Plasma biomarkers of brain injury in neonatal hypoxic-ischemic encephalopathy". In: *The Journal of pediatrics* 194 (2018), pp. 67–75.

[18] Martha Douglas-Escobar and Michael D Weiss. "Biomarkers of hypoxic-ischemic encephalopathy in newborns". In: *Frontiers in neurology* 3 (2012), p. 144.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas 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, pp. 234–241.

[20] Keelin Murphy, Niek E van der Aa, Simona Negro, Floris Groenendaal, Linda S de Vries, Max A Viergever, Geraldine B Boylan, Manon JNL Benders, and Ivana Išgum. "Automatic quantification of ischemic injury on diffusion-weighted MRI of neonatal hypoxic ischemic encephalopathy". In: *NeuroImage: Clinical* 14 (2017), pp. 222–232.

[21] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. "The multimodal brain tumor image segmentation benchmark (BRATS)". In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.

[22] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific data* 4.1 (2017), pp. 1–13.

[23] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge". In: *arXiv preprint arXiv:1811.02629* (2018).

[24] Oskar Maier, Bjoern H Menze, Janina Von der Gablentz, Levin Häni, Mattias P Heinrich, Matthias Liebrand, Stefan Winzeck, Abdul Basit, Paul Bentley, Liang Chen, et al. "ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI". In: *Medical image analysis* 35 (2017), pp. 250–269.

[25] Aaron Carass, Snehashis Roy, Amod Jog, Jennifer L Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button, James Nguyen, Ferran Prados, Carole H Sudre, et al. "Longitudinal multiple sclerosis lesion segmentation: resource and challenge". In: *NeuroImage* 148 (2017), pp. 77–102.

[26] Olivier Commowick, Audrey Istace, Michael Kain, Baptiste Laurent, Florent Leray, Mathieu Simon, Sorina Camarasu Pop, Pascal Girard, Roxana Ameli, Jean-Christophe Ferré, et al. "Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure". In: *Scientific reports* 8.1 (2018), p. 13650.

[27] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. "The medical segmentation decathlon". In: *Nature communications* 13.1 (2022), p. 4128.

[28] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211.

[29] Shawn N Murphy, Christopher Herrick, Yanbing Wang, Taowei David Wang, Darren Sack, Katherine P Andriole, Jesse Wei, Nathaniel Reynolds, Wendy Plesniak, Bruce R Rosen, et al. "High throughput tools to access images from clinical archives for research". In: *Journal of digital imaging* 28 (2015), pp. 194–204.

[30] Christoph Forman, Jens Wetzl, Carmel Hayes, and Michaela Schmidt. "Compressed sensing: a paradigm shift in MRI". In: *MAGNETOM Flash* 66 (2016), pp. 9–13.

[31] James M Provenzale, Luxia Liang, David DeLong, and Leonard E White. "Diffusion tensor imaging assessment of brain white matter maturation during the first postnatal year". In: *American Journal of Roentgenology* 189.2 (2007), pp. 476–486.

[32] Jeroen Dudink, Maarten Lequin, Carola van Pul, Jan Buijs, Nikk Conneman, Johannes van Goudoever, and Paul Govaert. "Fractional anisotropy in white matter tracts of very-low-birth-weight infants". In: *Pediatric radiology* 37 (2007), pp. 1216–1223.

[33] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. "N4ITK: improved N3 bias correction". In: *IEEE transactions on medical imaging* 29.6 (2010), pp. 1310–1320.

[34] Yangming Ou, Lilla Zöllei, Xiao Da, Kallirroi Retzepi, Shawn N Murphy, Elizabeth R Gerstner, Bruce R Rosen, P Ellen Grant, Jayashree Kalpathy-Cramer, and Randy L Gollub. "Field of view normalization in multi-site brain MRI". In: *Neuroinformatics* 16.3 (2018), pp. 431–444.

[35] Yangming Ou, Randy L Gollub, Kallirroi Retzepi, Nathaniel Reynolds, Rudolph Pienaar, Steve Pieper, Shawn N Murphy, P Ellen Grant, and Lilla Zöllei. "Brain extraction in pediatric ADC maps, toward characterizing neuro-development in multi-platform and multi-institution clinical images". In: *NeuroImage* 122 (2015), pp. 246–261.

[36] Yangming Ou, Lilla Zöllei, Kallirroi Retzepi, Victor Castro, Sara V Bates, Steve Pieper, Katherine P Andriole, Shawn N Murphy, Randy L Gollub, and Patricia Ellen Grant. "Using clinically acquired MRI to construct age-specific ADC atlases: Quantifying spatiotemporal ADC changes from birth to 6-year old". In: *Human brain mapping* 38.6 (2017), pp. 3052–3068.

[37] Yangming Ou, Aristeidis Sotiras, Nikos Paragios, and Christos Davatzikos. "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting". In: *Medical image analysis* 15.4 (2011), pp. 622–639.

[38] Yangming Ou, Hamed Akbari, Michel Bilello, Xiao Da, and Christos Davatzikos. "Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights". In: *IEEE transactions on medical imaging* 33.10 (2014), pp. 2039–2065.

[39] Yangming Ou, Aristeidis Sotiras, Nikos Paragios, and Christos Davatzikos. "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting". In: *Medical image analysis* 15.4 (2011), pp. 622–639.

[40] Anna LR Pinto, Yangming Ou, Mustafa Sahin, and P Ellen Grant. "Quantitative apparent diffusion coefficient mapping may predict seizure onset in children with Sturge-Weber syndrome". In: *Pediatric neurology* 84 (2018), pp. 32–38.

[41] Sheng He, Diana Pereira, Juan David Perez, Randy L Gollub, Shawn N Murphy, Sanjay Prabhu, Rudolph Pienaar, Richard L Robertson, P Ellen Grant, and Yangming Ou. "Multi-Channel attention-fusion neural network for brain age estimation: accuracy, generality, and interpretation with 16,705 healthy MRIs across lifespan". In: *Medical Image Analysis* 72 (2021), p. 102091.

[42] Sheng He, P Ellen Grant, and Yangming Ou. "Global-local transformer for brain age estimation". In: *IEEE transactions on medical imaging* 41.1 (2021), pp. 213–224.

[43] Jennifer E Khoury, Banu Ahtam, Michaela Sisitsky, Yangming Ou, Borjan Gagoski, Michelle Bosquet Enlow, Martin H Teicher, P Ellen Grant, and Karlen Lyons-Ruth. "Maternal childhood maltreatment is associated with lower infant gray matter volume and amygdala volume during the first two years of life". In: *Biological psychiatry global open science* 2.4 (2022), pp. 440–449.

[44] Martha Douglas-Escobar and Michael D Weiss. "Hypoxic-ischemic encephalopathy: a review for the clinician". In: *JAMA pediatrics* 169.4 (2015), pp. 397–403.

[45] Ruili Wei, Chaonan Wang, Fangping He, Lirong Hong, Jie Zhang, Wangxiao Bao, Fangxia Meng, and Benyan Luo. "Prediction of poor outcome after hypoxic-ischemic brain injury by diffusion-weighted imaging: A systematic review and meta-analysis". In: *Plos one* 14.12 (2019), e0226295.

[46] Susan Sotardi, Randy L Gollub, Sara V Bates, Rebecca Weiss, Shawn N Murphy, P Ellen Grant, and Yangming Ou. "Voxelwise and regional brain apparent diffusion coefficient changes on MRI from birth to 6 years of age". In: *Radiology* 298.2 (2021), p. 415.

[47] Yangming Ou, Lilla Zöllei, Kallirroi Retzepi, Victor Castro, Sara V Bates, Steve Pieper, Katherine P Andriole, Shawn N Murphy, Randy L Gollub, and Patricia Ellen Grant. "Using clinically acquired MRI to construct age-specific ADC atlases: Quantifying spatiotemporal ADC changes from birth to 6-year old". In: *Human brain mapping* 38.6 (2017), pp. 3052–3068.

[48] Lu Liu, Jelmer M Wolterink, Christoph Brune, and Raymond NJ Veldhuis. "Anatomy-aided deep learning for medical image segmentation: a review". In: *Physics in Medicine & Biology* 66.11 (2021), 11TR01.

[49] Toshiyuki Imanishi, Masaki Shimizu, Wakako Sumiya, Chika Kanno, Masayuki Kanno, Masami Kanno, and Ken Kawabata. "Brain injury following mild hypoxic-ischemic encephalopathy in neonates–Ten-year experience in a tertiary perinatal center". In: *Journal of Perinatology* (2022), pp. 1–7.

[50] Yi Li, Jessica L Wisnowski, Lina Chalak, Amit M Mathur, Robert C McKinstry, Genesis Licona, Dennis E Mayock, Taeun Chang, Krisa P Van Meurs, Tai-Wei Wu, et al. "Mild hypoxic-ischemic encephalopathy (HIE): Timing and pattern of MRI brain injury". In: *Pediatric research* (2022), pp. 1–6.

[51] Lina Chalak, Samantha Latremouille, Imran Mir, Pablo J Sánchez, and Guilherme Sant'Anna. "A review of the conundrum of mild hypoxic-ischemic encephalopathy: Current challenges and moving forward". In: *Early human development* 120 (2018), pp. 88–94.

[52] Abbot R Laptook, Seetha Shankaran, Jon E Tyson, Breda Munoz, Edward F Bell, Ronald N Goldberg, Nehal A Parikh, Namasivayam Ambalavanan, Claudia Pedroza, Athina Pappas, et al. "Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial". In: *Jama* 318.16 (2017), pp. 1550–1560.

[53] A James Barkovich, Beatrice Latal Hajnal, Daniel Vigneron, Augusto Sola, J Colin Partridge, Faith Allen, and Donna M Ferriero. "Prediction of neuromotor outcome in perinatal asphyxia: evaluation of MR scoring systems." In: *American Journal of Neuroradiology* 19.1 (1998), pp. 143–149.

[54] Shamik B Trivedi, Zachary A Vesoulis, Rakesh Rao, Steve M Liao, Joshua S Shimony, Robert C McKinstry, and Amit M Mathur. "A validated clinical MRI injury scoring system in neonatal hypoxic-ischemic encephalopathy". In: *Pediatric radiology* 47.11 (2017), pp. 1491–1499.

[55] Lishya Liauw, Gerda van Wezel-Meijler, Sylvia Veen, MA Van Buchem, and Jeroen van der Grond. "Do apparent diffusion coefficient measurements predict outcome in children with neonatal hypoxic-ischemic encephalopathy?" In: *American journal of neuroradiology* 30.2 (2009), pp. 264–270.

[56] Rod W Hunt, Jeffrey J Neil, Lee T Coleman, Michael J Kean, and Terrie E Inder. "Apparent diffusion coefficient in the posterior limb of the internal capsule predicts outcome after perinatal asphyxia". In: *Pediatrics* 114.4 (2004), pp. 999–1003.

[57] Niek E van der Aa, Manon JNL Benders, Koen L Vincken, Floris Groenendaal, and Linda S de Vries. "The course of apparent diffusion coefficient values following perinatal arterial ischemic stroke". In: *PLoS One* 8.2 (2013), e56784.

[58] Sheng He, Yanfang Feng, P Ellen Grant, and Yangming Ou. "Segmentation ability map: Interpret deep features for medical image segmentation". In: *Medical Image Analysis* 84 (2023), p. 102726.

[59] Supriya Kushwah, Ashok Kumar, Ashish Verma, Sriparna Basu, and Ashutosh Kumar. "Comparison of fractional anisotropy and apparent diffusion coefficient among hypoxic ischemic encephalopathy stages 1, 2, and 3 and with nonasphyxiated newborns in 18 areas of brain". In: *Indian Journal of Radiology and Imaging* 27.04 (2017), pp. 447–456.

[60] Jun Shibasaki, Tetsu Niwa, Aurélie Piedvache, Moyoko Tomiyasu, Naho Morisaki, Yuta Fujii, Katsuaki Toyoshima, and Noriko Aida. "Comparison of predictive values of magnetic resonance biomarkers based on scan timing in neonatal encephalopathy following therapeutic hypothermia". In: *The Journal of Pediatrics* 239 (2021), pp. 101–109.

[61] Fatema Al Amrani, Saskia Kwan, Guillaume Gilbert, Christine Saint-Martin, Michael Shevell, and Pia Wintermark. "Early imaging and adverse neurodevelopmental outcome in asphyxiated newborns treated with hypothermia". In: *Pediatric neurology* 73 (2017), pp. 20–27.

[62] Eva-Marie Heursen, Amaya Zuazo Ojeda, Isabel Benavente Fernández, Gema Jimenez Gómez, Rosalía Campuzano Fernández-Colima, José Paz-Expósito, and Simón Pedro Lubián López. "Prognostic value of the apparent diffusion coefficient in newborns with hypoxic-ischaemic encephalopathy treated with therapeutic hypothermia". In: *Neonatology* 112.1 (2017), pp. 67–72.