



# Deep Multimodal Learning From MRI and Clinical Data for Early Prediction of Neurodevelopmental Deficits in Very Preterm Infants

Lili He<sup>1,2,3\*</sup>, Hailong Li<sup>1,2</sup>, Ming Chen<sup>1,2,4</sup>, Jinghua Wang<sup>3</sup>, Mekibib Altaye<sup>5</sup>, Jonathan R. Dillman<sup>1,2,3</sup> and Nehal A. Parikh<sup>6,7</sup>

<sup>1</sup> Imaging Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, <sup>2</sup> Department of Radiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, <sup>3</sup> Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, OH, United States, <sup>4</sup> Department of Electronic Engineering and Computing Systems, University of Cincinnati, Cincinnati, OH, United States, <sup>5</sup> Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, <sup>6</sup> The Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, <sup>7</sup> Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States

## OPEN ACCESS

### Edited by:

Yalin Wang,  
Arizona State University, United States

### Reviewed by:

Qunxi Dong,  
Beijing Institute of Technology, China  
Yaser A. ElNakieb,  
University of Louisville, United States

### \*Correspondence:

Lili He  
lili.he@cchmc.org

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 04 August 2021

**Accepted:** 13 September 2021

**Published:** 05 October 2021

### Citation:

He L, Li H, Chen M, Wang J, Altaye M, Dillman JR and Parikh NA (2021) Deep Multimodal Learning From MRI and Clinical Data for Early Prediction of Neurodevelopmental Deficits in Very Preterm Infants. *Front. Neurosci.* 15:753033. doi: 10.3389/fnins.2021.753033

The prevalence of disabled survivors of prematurity has increased dramatically in the past 3 decades. These survivors, especially, very preterm infants (VPIs), born  $\leq 32$  weeks gestational age, are at high risk for neurodevelopmental impairments. Early and clinically effective personalized prediction of outcomes, which forms the basis for early treatment decisions, is urgently needed during the peak neuroplasticity window—the first couple of years after birth—for at-risk infants, when intervention is likely to be most effective. Advances in MRI enable the noninvasive visualization of infants' brains through acquired multimodal images, which are more informative than unimodal MRI data by providing complementary/supplementary depicting of brain tissue characteristics and pathology. Thus, analyzing quantitative multimodal MRI features affords unique opportunities to study early postnatal brain development and neurodevelopmental outcome prediction in VPIs. In this study, we investigated the predictive power of multimodal MRI data, including T2-weighted anatomical MRI, diffusion tensor imaging, resting-state functional MRI, and clinical data for the prediction of neurodevelopmental deficits. We hypothesize that integrating multimodal MRI and clinical data improves the prediction over using each individual data modality. Employing the aforementioned multimodal data, we proposed novel end-to-end deep multimodal models to predict neurodevelopmental (i.e., cognitive, language, and motor) deficits independently at 2 years corrected age. We found that the proposed models can predict cognitive, language, and motor deficits at 2 years corrected age with an accuracy of 88.4, 87.2, and 86.7%, respectively, significantly better than using individual data modalities. This current study can be considered as proof-of-concept. A larger study with external validation is important to validate our approach to further assess its clinical utility and overall generalizability.

**Keywords:** deep learning, neurodevelopment, very preterm infants, MRI, resting state functional MRI, diffusion tensor imaging, brain connectome, diffuse white matter abnormality

## INTRODUCTION

With the continuing high incidence of preterm births (about 380,000 in 2018) (Martin et al., 2019) and improving survival rates (exceeding 90%) (Blencowe et al., 2012) in the United States, the prevalence of disabled survivors of prematurity has increased dramatically. These survivors, especially, very preterm infants (VPIs), born  $\leq 32$  weeks gestational age (GA), are at high risk for cognitive deficits and other neurodevelopmental disorders, thereby increasing their risk for poor educational, health, and social outcomes (Jarjour, 2015). Efforts to target interventions to prevent and/or treat neurodevelopmental sequelae are hampered by our current inability to diagnose or predict risk of disabilities before the age of 3–5 years (Nordhov et al., 2010; Kwon et al., 2014). The imminent challenge lies in early identification of infants who are at the greatest risk for developing later disorders at an individual level. Early and clinically effective personalized prediction of outcomes, which forms the basis for early treatment decisions, is urgently needed during the peak neuroplasticity window—the first couple of years after birth—for at-risk infants, when intervention is likely to be most effective (Johnston, 2009).

Advances in MRI enable the noninvasive visualization of infants' brains through acquired multi-modal images. Research supports the findings that brain imaging features are modulated by genetic (Thompson et al., 2001), non-genetic biological (Hackman and Farah, 2009), and environmental (May, 2011) influences, and therefore show high variability among subjects. Such variability can potentially provide valuable information for personalized prognosis based on the characteristics of individual patients (Valizadeh et al., 2018). Brain anatomical features have been recently extended to the prognostication of neurodevelopmental impairments (cognitive, motor, working memory, and language), autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD) (Boardman et al., 2010; Thompson et al., 2014; Chaddad et al., 2017). We have externally validated our findings (He and Parikh, 2013; Li et al., 2019; Parikh et al., 2020) that objectively-diagnosed diffuse white matter abnormality (DWMA) at term equivalent age is an independent predictor of cognitive and language development in VPIs. In addition, brain connectivity patterns are formed during early brain development and reshaped in cases of prematurity or perinatal brain injury (Cao et al., 2017). Brain connectome studies have revealed microstructural alterations in cognition and motor tracts that correlate with poorer cognitive and motor performance (Thompson et al., 2014; Rogers et al., 2016). Atypical functional connectivity has been reported in children who develop adverse cognitive, language and motor outcomes (He and Parikh, 2015; Gozdas et al., 2018; He et al., 2020). Multimodal MRI data are more informative than unimodal MRI data by providing complementary/supplementary depicting of how brain tissue characteristics and their pathology information are segregated and integrated. Therefore, accurately analyzing quantitative multimodal MRI features affords unique opportunities to study early postnatal brain development and neurodevelopmental outcome prediction in preterm infants (Thompson et al., 2016). Through this, we may gain a better

understanding of how an individual brain's organizational changes influence cognitive, language, and motor functions.

Although, it is easy to understand, how to endow machines with capabilities to perceive patients through comprehensive information from multiple imaging or other data modalities is still an open question. The feature representations from different modalities originally locate in unequal subspaces, resulting that similar feature representations may be associated with completely different semantics. Therefore, the biggest challenge is how to project heterogeneous features into a common space, where the multimodal data with similar semantics will be represented by similar features (Rasiwasia et al., 2010; Guo et al., 2019). In the computer vision domain, studies have been conducted to address this problem in various applications, such as, video description and classification (Liu et al., 2016), event detection (Wu et al., 2014), cross-modal retrieval and translation (Qi and Peng, 2018; Wu et al., 2018), image caption (Xu et al., 2015), and text-to-image synthesis (Reed et al., 2016). In light of these existing works, and with recent advances in deep learning techniques (Hjelm et al., 2014; Plis et al., 2014; Mostapha and Styner, 2019), we propose to encode each unimodal representation, and then fuse the encoded unimodal features.

Unlike most published studies that describe unimodal MRI data (Kawahara et al., 2017; Moeskops et al., 2017; Girault et al., 2019; He et al., 2020; Saha et al., 2020), in this paper, we employed multimodal MRI and proposed deep multimodal learning models. We hypothesize that integrating multimodal MRI and clinical data improves early prediction of cognitive, language, and motor deficits independently, at 2 years corrected age in VPIs over using each individual data modality. By doing so, the proposed prediction model is capable of analyzing different types of inputs by fusing different neural networks. Specifically, the different model inputs, which were all collected at term-equivalent age, include: (1) structural brain connectome data from diffusion tensor imaging (DTI); (2) functional brain connectome data from resting-state functional MRI (rs-fMRI) connectome data; (3) DWMA quantified from anatomical T2-weighted images; and (4) perinatal clinical data. The fusion technique used here is a concatenation of the four encoded feature vectors, which is then used as an input to fully-connected layers before the network outputs its prediction. The resulting classification system is a deep multimodal learning model, an automated prognostic system that uses four types of data as inputs to determine at term-equivalent age whether or not an individual VPI is at high risk of developing moderate or more severe cognitive, language, and/or motor deficits and to predict individual standardized neurodevelopmental scores (on the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley III) (Bayley, 2009) Cognitive, Language, and Motor subtest scores) at 2 years corrected age.

The main contributions of our work are highlighted as follows: (1) We proposed end-to-end deep multimodal learning models that incorporate features from multimodal MRI (anatomical, DTI, and rs-fMRI) and clinical data; (2) We demonstrated that the application of deep multimodal learning to analyze high-dimensional objectively-quantified anatomical and connectome features may detect brain structural and functional abnormalities

and tissue pathology that are not readily visible to the naked eye, thereby facilitating risk stratification; (3) We unwrapped and identified discriminative MRI and clinical features used by the proposed models to make predictions. Such discriminative feature identification will generate greater trust in the prognostic models and enhanced pathophysiological understanding.

## METHODS AND MATERIALS

### Subjects and MRI Acquisition

The Institutional Review Boards of the Nationwide Children's Hospital (NCH) and Cincinnati Children's Hospital Medical Center (CCHMC) approved this study, and written parental informed consent was obtained for every subject. This study has been carried out in accordance with The Code of Ethics of the World Medical Association. This study included 261 prospectively recruited VPIs from five Cincinnati Ohio neonatal intensive care units (NICUs) as cohort I (for unsupervised model pre-training), and 108 VPIs from four Columbus area/Central Ohio NICUs as cohort II (for supervised model fine-tuning). All subjects were scanned during natural sleep without the use of any sedation after being fed and swaddled. Infants with congenital structural central nervous system anomalies (e.g., Dandy-Walker, encephalocele, diffuse calcifications, and meningomyelocele) or congenital chromosomal abnormalities known to be associated with neurodevelopmental impairments were excluded.

Subjects in the Cohort I were scanned at 39–44 weeks postmenstrual age (PMA) on a 3T MRI scanner (Ingenia, Philips Healthcare, Best, The Netherlands) at CCHMC using a 32-channel head coil. Anatomical scans were conducted with a 2D T2-weighted fast spin-echo sequence. Functional MRI data were conducted using multi-band rs-fMRI (multi-band factor = 3). Diffusion MRI data were collected using single-shot echo planar imaging (EPI). Detailed acquisition parameters are listed in **Supplementary Table 1**.

All cohort II subjects were scanned at 38–43 weeks PMA on a 3T MRI scanner (Skyra; Siemens Healthcare) at NCH using a 32-channel head coil. Anatomical scans were conducted with a 2D T2-weighted fast spin-echo sequence. Functional MRI data were collected using single-band/multi-band rs-fMRI (multi-band factor = 3). Diffusion MRI data were collected using single-shot EPI. Detailed acquisition parameters are also listed in **Supplementary Table 1**.

### Clinical Features and Neurodevelopmental Assessments

For each VPI, 72 a priori defined and prospectively collected perinatal clinical features were retrieved (**Supplementary Table 2**). Clinical features related to five overarching domains, including: (1) maternal demographics (e.g., mothers age, gravida, parity, mother's highest educational level, etc.); (2) pregnancy complications (e.g., diabetes, hypertension, hypothyroidism, etc.); (3) labor and delivery (e.g., rupture of membranes, antenatal steroids, magnesium administration, etc.); (4) neonatal information at birth (e.g., sex, gestational age, birth weight, etc.); and (5) medical history (e.g.,

oxygen or positive pressure support, surfactant administration, pneumothorax, sepsis, bronchopulmonary dysplasia, etc.).

The Bayley III Cognitive, Language, and Motor subtest scores [each standardized on a scale of 40–160, with a mean of 100 and standard deviation (SD) of 15] served as the primary neurodevelopmental outcome measures. We dichotomized the VPIs using Bayley-III score of 90 into those high-risk ( $\leq 90$ ) vs. low-risk ( $> 90$ ) for neurodevelopmental deficits.

### DWMA Quantification

We quantified DWMA using our published objective algorithm (He and Parikh, 2013). Briefly, brain tissue segmentation (white matter, gray matter, and cerebrospinal fluid) was achieved by unified segmentation on T2-weighted images with spatial priors obtained from a neonatal probabilistic atlas (Shi et al., 2011). We considered voxels with signal intensity values greater than  $\alpha$  standard deviation above the mean of cerebral (white + gray matter) tissues to be DWMA. Volume of DWMA was calculated as the product of voxel volume and total number of voxels in the detected DWMA region. We determined the normalized volume of DWMA by dividing DWMA volume by total cerebral white matter volume. The optimal  $\alpha$  may be different for different cohort MRI data acquired with different imaging protocols (He and Parikh, 2013, 2015; Li et al., 2019; Parikh et al., 2020). Instead of determining one single optimal  $\alpha$  value, in this work, to take advantage of the strength of feature integration, we defined a DWMA feature vector which contained a series of DWMA volumes that were obtained by varying the threshold  $\alpha$  from 1.4 to 2.0 with increment of 0.1. To control inter-subject variability, we also include the volume of white matter, gray matter, and CSF as confounders into the DWMA feature vector.

### Structural Connectome Quantification

We preprocessed DTI data with a pipeline involving skull stripping, registration, head motion, and eddy current artifacts correction using FMRIB Software Library (FSL, Oxford University, UK) (Woolrich et al., 2009). We conducted diffusion tensor reconstruction based on a linear least-square fitting algorithm and brain fiber tracking based on a deterministic tracking algorithm in the subject's native space using Diffusion Toolkit/TrackVis (Wang et al., 2007). We harmonized fractional anisotropy maps using a batch-effect correction algorithm ComBat (Fortin et al., 2017) to remove undesirable variabilities caused by different acquisition parameters. The brain was parcellated into 90 regions of interest (ROIs) according to a neonatal anatomical template (Shi et al., 2011), forming the nodes of the individual structural networks. Structural connectivity map (i.e.,  $90 \times 90$  network adjacency matrix symmetric about the diagonal), were constructed using the UCLA Multimodal Connectivity Package (Bassett et al., 2011). Each entry in the structural connectome map represents the brain structural connectivity between each pair of ROIs, which was calculated as the mean fractional anisotropy of each voxel intersecting the tract and then averaged over all tracts between the two nodes.

## Functional Connectome Quantification

We performed rs-fMRI preprocessing using previously validated pipelines (Pogribna et al., 2014; He and Parikh, 2016), to (1) Reorient all acquired scans with anterior commissure (AC)—posterior commissure (PC) line; (2) Remove non-brain parts of the image; (3) Correct motion artifact by aligning each time point's frame to the middle frame, and estimate corresponding six motion parameters [three translation (displacement) and three rotation parameters]; (4) Register both rs-fMRI and structural T2-weighted images to be in the same “standard space” [a neonatal brain atlas (Shi et al., 2011)]; (5) Regress out the mean time courses of cerebral white matter, ventricles, and whole brain and their derivatives; as well as six motion parameters and their derivatives and squares (Power et al., 2014); (6) Improve signal-to-noise ratio and ameliorate the effects of functional misalignments across subjects (Lowe and Sorenson, 1997) using spatial smoothing with isotropic Gaussian filter with 6 mm kernel; and (7) remove the lowest and highest temporal drifts in the data *via* band-pass filtering ( $0.008 < f < 0.09$  Hz; Hallquist et al., 2013). We then parcellated the brain into 223 ROIs according to a neonatal functional template (Shi et al., 2017), forming the nodes of the individual brain functional networks. We extracted rs-fMRI time series from each ROI, then computed the functional connectivity as the correlation between the time series of each pair of ROIs. This resulted in a functional connectome map (i.e.,  $223 \times 223$  network adjacency matrix symmetric about the diagonal). All above operations were conducted using FMRIB Software Library (FSL, Oxford University, UK), Statistical Parametric Mapping software (SPM, University College London, UK; Friston, 1994) and functional connectivity toolbox (CONN) (Whitfield-Gabrieli and Nieto-Castanon, 2012). We also conducted connectome map harmonization using the ComBat algorithm (Fortin et al., 2017).

## Data Augmentation and Balancing

We conducted data augmentation and balancing on the training data to enable a robust model training. A challenge in the proposed supervised model training is the relatively small number of infants at high-risk compared to those at low-risk. Imbalanced datasets can severely affect the model's learning ability (Haixiang et al., 2017). In such cases, the deep learning models may become majority class classifiers, i.e., they fail to learn the concepts of the minority class. To overcome this challenge, we employed a data balancing and augmentation method (Kawahara et al., 2017), which uses neighborhood samples to create artificial minority samples. By synthetically generating more samples of the minority class, the classifiers are able to broaden their decision regions for the minority class. Specifically, similar to a prior work (Kawahara et al., 2017), we first categorized supervised training dataset into five bins according to a VPI's Bayley-III subtest score ( $<70$ ,  $70-79$ ,  $80-89$ ,  $90-100$ , and  $>100$ ). We randomly selected a sample (i.e., functional, or structural connectivity data) in a bin with the fewest samples and searched for  $k$  nearest neighbors for the given sample based on Euclidean distance. Assuming that the selected sample is  $x_0$ , and its associated neighbors are  $[x_1, \dots, x_i, \dots, x_k]$ , a synthetic data  $x_s$  is generated by:  $x_s = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_i x_i +$

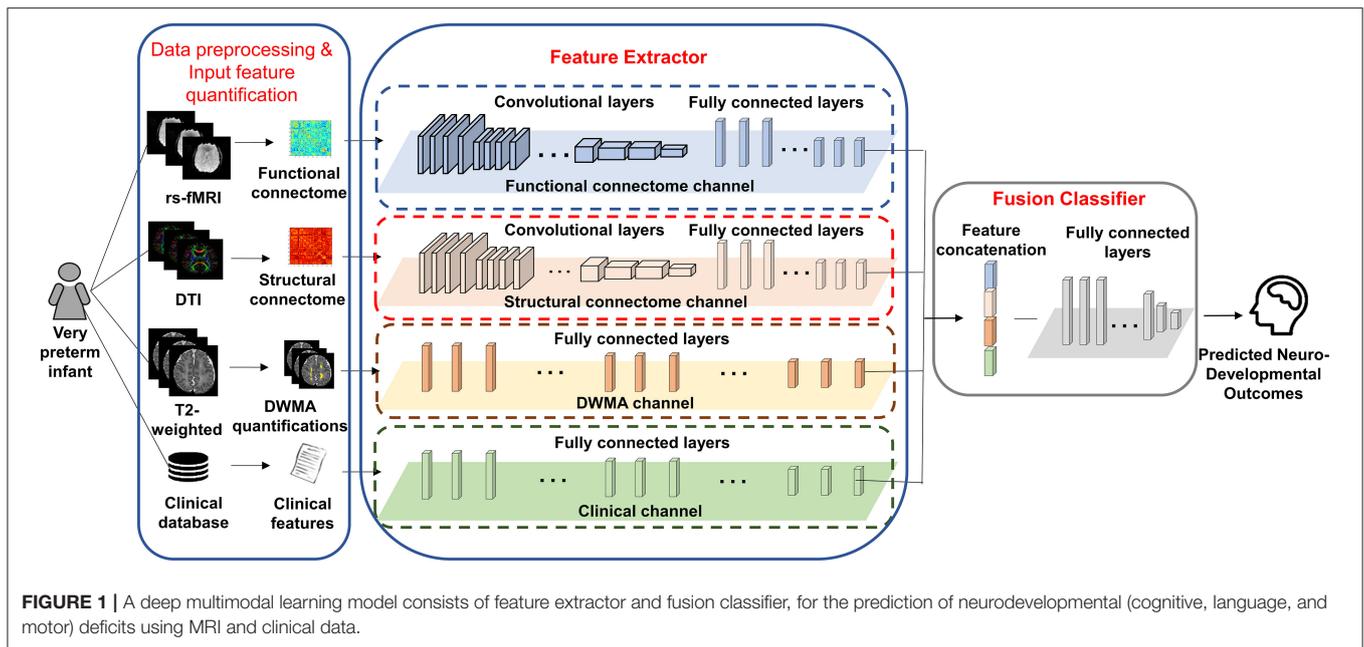
$\dots + \beta_k x_k$ , where  $\beta_i$  is a random weight, and  $\sum_{i=0}^k \beta_i = 1$ . The corresponding Bayley-III score  $y_s$  was generated in the same way. We repeatedly generated synthetic samples for each bin until the numbers of training samples in all bins were equal. This process was also repeated until the number of training samples reached 10 times that of the original training dataset. Importantly, the synthetic data were only used for model training, but not for testing.

## Model Design

We proposed deep multimodal learning models for the early prediction of cognitive, language, and motor deficits using multimodal MRI and clinical data (Figure 1). We have presented how imaging and clinical data were acquired and preprocessed, as well as how multimodal MRI features were quantified in subsections (Subjects and MRI Acquisition, Clinical Features and Neurodevelopmental Assessments, DWMA Quantification, Structural Connectome Quantification, Functional Connectome Quantification, and Data Augmentation and Balancing). Each of our proposed models contain a feature extractor and a fusion classifier. The feature extractor has four parallel channels to extract discriminative high-level functional and structural connectivity, DWMA, and clinical features out of high-dimensional input data, respectively. Both functional and structural connectivity channels have the same network architecture. It consists of 16 convolutional layers and 5 pooling layers adopted from the pre-trained VGG-19 model (Simonyan and Zisserman, 2014), followed by fully connected blocks. Since the feature dimensions of the DWMA and clinical data are not high, both DWMA and clinical channels only consist of fully connected blocks, without pre-trained VGG-19 layers for the feature dimensionality reduction. Each fully connected block contains a fully connected layer, a batch normalization, and a dropout layer. The dropout layer is a regularization technique that randomly selects a certain ratio of neurons and ignores them during training (Srivastava et al., 2014). The “dropped-out” neurons do not contribute to the feedforward process, and the weights of these neurons are not updated in backpropagation. Dropout regularization helps avoid model overfitting. Batch normalization solves the internal covariate shift problem (Ioffe and Szegedy, 2015). Similar to feature scaling, batch normalization works to adjust, and scale hidden unit shifts across hidden layers. Batch normalization also speeds up the training process when handling a large number of features. Finally, we design a fusion classifier to integrate the discriminative information from all extracted high-level imaging and clinical features using a fully connected layer with one output neuron. We conduct the outcome classification using a softmax function and outcome regression using a linear function.

## Model Training and Optimization

Deep learning models generally require training on large datasets to achieve good performance while our annotated dataset for the target tasks (i.e., prediction of cognitive, language, and motor deficits) is relatively small. To address this issue, we utilized both supervised and unsupervised transfer learning approaches. In particular, the VGG-19 (Simonyan and Zisserman, 2014)



**FIGURE 1 |** A deep multimodal learning model consists of feature extractor and fusion classifier, for the prediction of neurodevelopmental (cognitive, language, and motor) deficits using MRI and clinical data.

layers described above were pretrained with supervision using ImageNet database (~1.2 million images). The weights of these layers were fixed and reused in our model. The weights of all other neural network layers were first pretrained without supervision using a relatively large unannotated VPI data from cohort I. These weights were finally retrained and fine-tuned in a supervised fashion using annotated VPI data from cohort II for outcome classification/regression. The mechanism behind this rationale is that we can repurpose models developed for other tasks utilizing a large dataset to ultimately improve the performance and generalizability of our proposed models as well as decrease the amount of data needed for model training.

Specifically, given  $m$  training samples in the cohort I,  $[\mathbf{X}_f^i, \mathbf{X}_s^i, \mathbf{x}_d^i, \mathbf{x}_c^i]$ ,  $i \in [1, m]$  are the input data of the  $i$ -th sample without label, where  $\mathbf{X}_f^i$  is a two-dimensional adjacency matrix (i.e.,  $223 \times 223$ ) of functional connectivity;  $\mathbf{X}_s^i$  is a two-dimensional adjacency matrix (i.e.,  $90 \times 90$ ) of structural connectivity;  $\mathbf{x}_d^i$  is the one-dimensional vector (i.e.,  $1 \times 11$ ) of DWMA measures; and  $\mathbf{x}_c^i$  is a one-dimensional vector (i.e.,  $1 \times 72$ ) of clinical data. As mentioned above, we first utilized pretrained VGG-19 layers to extract high-level morphological features of adjacency matrix from both functional and structural connectivity. The outputs of VGG-19 layers are flattened as one-dimensional vectors (i.e.,  $1 \times k$ ) and denoted by  $\mathcal{H}(\mathbf{X}_f^i)$  and  $\mathcal{H}(\mathbf{X}_s^i)$ .

Next, to mitigate the issue of mismatch between ImageNet database and the small annotated VPI dataset in cohort II, we continued to perform an unsupervised transfer learning using the relatively large unannotated VPI dataset from cohort I. Except for VGG-19 layers, we pretrained the weights of all other neural network layers of both functional and structural connectivity channels without supervision. We pretrained the fully connected

layers of both functional and structural connectivity channels using an unsupervised learning strategy. We constructed a stacked sparse autoencoder (SSAE) for the fully connected layers. A rectified linear unit (ReLU) activation function was used in hidden nodes, and a sigmoid unit was chosen in the output layer. For each brain connectivity channel, we minimized the mean squared error loss function:

$$\mathcal{L} = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k \left( \mathcal{H}(\mathbf{X}^i)^j - \hat{\mathcal{H}}(\mathbf{X}^i)^j \right)^2$$

where  $\hat{\mathcal{H}}(\mathbf{X}^i)^j$  is the reconstructed functional or structural input  $\mathcal{H}(\mathbf{X}^i)^j$  from  $j$ -th neuron of the SSAE. A mini-batch Adam algorithm (Kingma and Ba, 2014) was selected to minimize the loss function. The learning rate was selected from empirical values [0.001, 0.01, 0.1, and 0.5]. Batch size was chosen using (Hackman and Farah, 2009; Johnston, 2009; Nordhov et al., 2010; Blencowe et al., 2012). Total number of epochs was 50. These hyperparameters were optimized based on validation data during model training/validation before model testing.

With these pretrained fully connected layers, we continued to retrain and fine-tune the whole model using a supervised training strategy using annotated VPI data from cohort II. Assume that there are  $n$  training samples in cohort II, and  $[\mathbf{X}_f^i, \mathbf{X}_s^i, \mathbf{x}_d^i, \mathbf{x}_c^i]$ ,  $i \in [1, n]$  are the input data of the  $i$ -th sample with label/score  $y^i$ ,  $i \in [1, n]$  (i.e., high risk vs. low risk of developing cognitive, language, or motor deficits). For the classification task, we fine-tuned the fully connected layers and fusion classifier of the model by minimizing cross-entropy loss

function as:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n y^i \log \left( p \left( y^i \mid \mathcal{H} \left( \mathbf{X}_f^i \right), \mathcal{H} \left( \mathbf{X}_s^i \right), \mathbf{x}_d^i, \mathbf{x}_c^i \right) \right) \\ + (1 - y^i) \log \left( 1 - p \left( y^i \mid \mathcal{H} \left( \mathbf{X}_f^i \right), \mathcal{H} \left( \mathbf{X}_s^i \right), \mathbf{x}_d^i, \mathbf{x}_c^i \right) \right)$$

where  $p \left( y^i \mid \mathcal{H} \left( \mathbf{X}_f^i \right), \mathcal{H} \left( \mathbf{X}_s^i \right), \mathbf{x}_d^i, \mathbf{x}_c^i \right)$  is the output of the fusion classifier, i.e., the probability of subject  $i$  being classified as the label  $y^i$ . For the score regression task, we applied a linear unit at the end of the model and optimized the mean absolute error loss function as follows:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n \left| y^i - \hat{y}^i \left( \mathcal{H} \left( \mathbf{X}_f^i \right), \mathcal{H} \left( \mathbf{X}_s^i \right), \mathbf{x}_d^i, \mathbf{x}_c^i \right) \right|$$

where  $\hat{y}^i \left( \mathcal{H} \left( \mathbf{X}_f^i \right), \mathcal{H} \left( \mathbf{X}_s^i \right), \mathbf{x}_d^i, \mathbf{x}_c^i \right)$  is the predicted output of the linear unit of the model, i.e., the predicted score. The mini-batch Adam algorithm was also used in the supervised learning. Training hyperparameters are listed in **Table 2**. To accelerate the model convergence, we applied an adaptive gradient update decay parameter (e.g., learning rate/maximal epoch). We used an early stop mechanism, which would cease the optimization process when multiple consecutive epochs returned the same validation loss errors.

With the fixed optimized pre-trained VGG19, our model architecture optimization focuses on the determination of the optimal number of fully connected layers and the optimal number of neurons at each layer. During the model training and validation, we tried the numbers of layers with empirical values from 1 to 4 in increments of 1; and we tried the numbers of neurons at each layer with empirical values from:  $2^n$ ,  $n \in [3, 4, 5, 6]$ . For each architecture setting, we ran 2-fold validations multiple times. According to the optimal validation performance, we set the optimal modal architecture (**Table 1**). The final training hyperparameters are listed in **Table 2**.

The proposed model development was implemented using Python 3.7.4, Keras (version: 2.1.6) with TensorFlow (version 1.14) backend on a computer workstation (256 GB RAM, 2 GPUs, Nvidia GTX1080 Ti).

## Most Discriminative Feature Identification

To unravel and illuminate the proposed deep multimodal learning models' predictive feature identification process and to generate greater trust in the models, we first adopted a feature ranking approach (Olden and Jackson, 2002) for one dimensional input of deep learning models to identify the most predictive clinical and DWMA risk factors. Specifically, we calculated the partial derivatives of the softmax output with respect to the clinical and DWMA features. For the softmax output (i.e., neurodevelopmental deficit)  $s$ , the partial derivatives  $\frac{\partial s}{\partial f_i^c}$  and  $\frac{\partial s}{\partial f_j^d}$ , where  $f_i^c$  is the  $i$ th clinical feature and  $f_j^d$  is the  $j$ th DWMA features, are computed for individual clinical and DWMA features. A higher absolute value of the partial derivative of  $\frac{\partial s}{\partial f_i^c}$  and  $\frac{\partial s}{\partial f_j^d}$  indicates a higher level of the importance for neurodevelopmental deficit prediction  $s$ .

We then implemented gradient-weighted class activation mapping (Grad-CAM) algorithm (Selvaraju et al., 2017), which was designed for two dimensional image input of deep learning models, to highlight both discriminative structural and functional brain connectivity in brain connectome maps (i.e., adjacency matrices). The Grad-CAM produces a coarse localization map highlighting predictive brain connectivities in the adjacency matrix by using gradient information of the last convolutional layer of the structural and functional channels (refer to **Figure 1** and **Table 1**). Specifically, we first computed the gradient of the softmax output  $s$  respect to the  $k$ th 2D feature map  $A$  of the last convolutional layer by  $\frac{\partial s}{\partial A_{ij}^k}$ , where  $i, j \in [1, m]$ , and  $m$  is the size of feature maps. Then, we obtained the weights of feature maps as  $\alpha_k = GAP \left( \frac{\partial s}{\partial A_{ij}^k} \right)$ , where  $GAP(*)$  is the global average pooling function. The heatmap of Grad-CAM was obtained by calculating the ReLU activation of the weighted combination of feature maps as:  $H = ReLU \left( \sum_k \alpha_k A^k \right)$ . The heatmap  $H$  was then normalized to  $[0, 1]$  and rescaled to the same size as adjacency matrices of structural and functional connectome. A higher value within  $H$  indicates a higher level of the importance for neurodevelopmental deficit prediction  $s$ .

## Model Validation

To evaluate the performance of the risk stratification (i.e., two-class classification), we calculated balanced accuracy, sensitivity, specificity, and area under the receiver operator characteristics curve (AUC). To evaluate the performance of the Bayley III score prediction (i.e., regression), we reported Pearson's correlation coefficient ( $r$ ), mean absolute error (MAE) and standard deviation of absolute error (SD of AE). We conducted nested five-fold cross-validation. In each iteration, the entire cohort II was divided into training data (60%), validation data (20%), and testing data (20%). Model optimization was conducted based on validating data without seeing testing data. We conducted this process for five iterations until all the cohort had been tested once. We then computed the performance across all five iterations. To test the reproducibility of the model, we repeated such five-fold cross-validation experiment 50 times and reported mean and standard deviation (SD).

## Statistical Analysis

Continuous demographic data and model performance metrics (described in the section Model Validation) were summarized as means and SDs, and categorical demographic data were summarized as counts and percentages. The two-sided Student's  $t$ -test (continuous data) and Chi-squared test (categorical data) were used to assess demographic characteristic differences between groups. The two-sided Student's  $t$ -test was also utilized to compare the model performances of using different feature sets. A  $p < 0.05$  was considered statistically significant. Analyses were performed with the statistical package of Matlab 2019b (MathWorks, Natick MA, United States).

**TABLE 1** | Architecture of the proposed deep multimodal learning model.

Layer index	Functional connectivity channel	Structural connectivity channel	DWMA channel	Clinical channel				
(1)	Conv2D, ReLU (fixed)	64@3×3	Conv2D, ReLU (fixed)	64@3×3	FC (trainable)	64	FC (trainable)	64
(2)	Conv2D, ReLU (fixed)	64@3×3	Conv2D, ReLU (fixed)	64@3×3	Batch Norm.	N/A	Batch Norm.	N/A
(3)	Maxpooling 2D	2	Maxpooling 2D	2	Dropout	0.2	Dropout	0.2
(4)	Conv2D, ReLU (fixed)	128@3×3	Conv2D, ReLU (fixed)	128@3×3	FC (trainable)	16	FC (trainable)	16
(5)	Conv2D, ReLU (fixed)	128@3×3	Conv2D, ReLU (fixed)	128@3×3	Batch Norm.	N/A	Batch Norm.	N/A
(6)	Maxpooling 2D	2	Maxpooling 2D	2	Dropout	0.2	Dropout	0.2
(7)	Conv2D, ReLU (fixed)	256@3×3	Conv2D, ReLU (fixed)	256@3×3				
(8)	Conv2D, ReLU (fixed)	256@3×3	Conv2D, ReLU (fixed)	256@3×3				
(9)	Conv2D, ReLU (fixed)	256@3×3	Conv2D, ReLU (fixed)	256@3×3				
(10)	Conv2D, ReLU (fixed)	256@3×3	Conv2D, ReLU (fixed)	256@3×3				
(11)	Maxpooling 2D	2	Maxpooling 2D	2				
(12)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(13)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(14)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(15)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(16)	Maxpooling 2D	2	Maxpooling 2D	2				
(17)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(18)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(19)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(20)	Conv2D, ReLU (fixed)	512@3×3	Conv2D, ReLU (fixed)	512@3×3				
(21)	Maxpooling 2D	2	Maxpooling 2D	2				
(22)	FC (trainable)	64	FC (trainable)	64				
(23)	Batch Norm.	N/A	Batch Norm.	N/A				
(24)	Dropout	0.2	Dropout	0.2				
(25)	FC (trainable)	16	FC (trainable)	16				
(26)	Batch Norm.	N/A	Batch Norm.	N/A				
(27)	Dropout	0.2	Dropout	0.2				
<b>Fusion classifier</b>								
(28)		FC (trainable)				8		
(29)		Batch Norm.				N/A		
(30)		Softmax				2		

Conv2D, 2D convolutional layer; Maxpooling 2D, 2D Maxpooling layer; Batch Norm. Batch Normalization layer; FC, Fully connected layer.

## RESULTS

### Subjects

After data quality control, excluding the data with largely incomplete brain coverage, high movement peaks, ghosting, incomplete imaging scans, and other scanner artifacts, we included 257 of 261 VPIs (mean (SD) GA at birth 29.3 (2.5) weeks; PMA at scan 42.7 (1.3) weeks; 111 (43.2%) male) without Bayley III assessments (cohort I), and 72 of 108 VPIs (mean (SD) GA at birth 28.3 (2.4) weeks; PMA at scan 40.3 (0.5) weeks; 41 (56.9%) male) with Bayley III assessments (cohort II). For all three neurodevelopmental (cognitive language, and motor) deficits prediction tasks, PMA was not significantly different between high-risk and low-risk groups. As expected, GA and birth weight were significantly different between the high-risk and low-risk groups. Additional demographic data for cohort II subjects with neurodevelopmental assessments at 2 years corrected age is listed in **Table 3**.

**TABLE 2** | Hyperparameters for the unsupervised and supervised training.

Hyperparameters	Value
High-risk neurodevelopmental deficits	>0
Low-risk neurodevelopmental deficits	0
Batch size	4
Dropout rate	0.2
Total number of epochs	50
Learning rate	0.01

### Cognitive Deficit Prediction

We tested the model performance of classifying VPIs into high- vs. low-risk group and predicting actual Bayley III Cognitive scores (i.e., continuous scale) using only clinical, functional connectome, structural connectome, and DWMA data alone; and then using combined features. As shown in **Table 4**, our model

**TABLE 3** | Demographic information of cohort II subjects with neurodevelopmental assessments at 2 years corrected age.

	Cognitive deficit		p-value
	Low risk (N = 47)	High risk (N = 25)	
PMA at early MRI in weeks; median (range)	40.3 (39.4–42.0)	40.1 (39.3–41.4)	0.5
GA in weeks; median (range)	29.1 (24.0–31.9)	28.0 (23.7–31.4)	0.04
Birth weight in grams; median (range)	1165 (570–2,340)	855 (510–1,755)	0.02
Sex; number of males (percentage)	27 (57.4%)	14 (56.0%)	0.906
	Language deficit		p-value
	Low risk (N = 47)	High risk (N = 25)	
PMA at early MRI in weeks; median (range)	40.3 (39.4–42.0)	40.1 (39.3–41.3)	0.14
GA in weeks; median (range)	29.7 (24.0–31.8)	26.8 (23.7–31.0)	0.0004
Birth weight in grams; median (range)	1285 (570–2,340)	855 (510–1,400)	<0.0001
Sex; number of males (percentage)	26 (55.3%)	15 (60.0%)	0.7025
	Motor deficit		p-value
	Low risk (N = 30)	High risk (N = 42)	
PMA at early MRI in weeks; median (range)	40.1 (39.4–41.3)	40.3 (39.3–42.0)	0.1
GA in weeks; median (range)	29.9 (25.1–31.9)	27.7 (23.7–31.4)	0.0003
Birth weight in grams; median (range)	1325 (715–1,900)	878 (510–2,340)	0.0004
Sex; number of males (percentage)	18 (60.0%)	23 (54.7%)	0.6581

PMA, postmenstrual age; GA, gestational age.

was able to correctly identify high-risk infants for cognitive deficits with a mean (SD) AUC of  $0.87 \pm 0.05$  and the Pearson's correlation coefficient  $r$  between the predicted and actual Bayley III Cognitive scores of  $0.62 \pm 0.04$  ( $p < 0.0001$ ) using the combined clinical and multi-modal MRI data. This was significantly greater than individually using only, (1) clinical data [AUC =  $0.74 \pm 0.05$  ( $p < 0.0001$ ) and  $r = 0.34 \pm 0.06$  ( $p < 0.0001$ )]; (2) functional connectome data [AUC =  $0.74 \pm 0.05$  ( $p < 0.0001$ ) and  $r = 0.34 \pm 0.07$  ( $p < 0.0001$ )]; (3) structural connectome data [AUC =  $0.81 \pm 0.06$  ( $p < 0.0001$ ) and  $r = 0.44 \pm 0.05$  ( $p < 0.0001$ )]; and (4) DWMA data [AUC =  $0.74 \pm 0.05$  ( $p < 0.0001$ ) and  $r = 0.39 \pm 0.04$  ( $p < 0.0001$ )]. These support our hypothesis that integrating multimodal MRI and clinical data

improves early prediction of cognitive deficits at 2 years corrected age in VPIs over using individual data modalities.

## Language Deficit Prediction

We next evaluated the model performance for language deficit risk stratification and Bayley III Language score prediction using individual and combined feature sets (Table 5). The model using the functional connectome alone achieved the lowest balanced accuracy of  $74.8 \pm 3.9\%$ , while the one using DWMA data alone had the lowest Pearson's correlation coefficient  $r$  of  $0.39 \pm 0.06$ . The deep multimodal learning model using combined features achieved the highest performance for risk stratification with a balanced accuracy of  $87.2 \pm 5.3\%$  and AUC of  $0.85 \pm 0.04$ . These were significantly higher than the second highest balanced accuracy of  $78.4 \pm 4.2\%$  ( $p < 0.0001$ ) using DWMA alone, and the second highest AUC of  $0.78 \pm 0.04$  ( $p < 0.0001$ ) using clinical features alone. The deep multimodal learning model achieved a Pearson's correlation coefficient  $r$  of  $0.63 \pm 0.04$  between the predicted and actual Bayley III language scores, significantly higher than the one using functional connectome ( $p < 0.0001$ ), structural connectome ( $p < 0.0001$ ), clinical data ( $p < 0.0001$ ), and DWMA data ( $p < 0.0001$ ). The results support our hypothesis that integrating multimodal MRI and clinical data improves early prediction of language deficits at 2 years corrected age in VPIs over using individual data modalities.

## Motor Deficit Prediction

Table 6 demonstrates the model performance for classifying high- vs. low-risk motor deficit group and predicting actual Bayley III Motor scores using individual and combined feature sets. The model using combined features was able to correctly identify high-risk VPIs for motor deficits with an AUC of  $0.85 \pm 0.06$ , significantly better than using functional connectome ( $0.71 \pm 0.05$ ;  $p < 0.0001$ ), structural connectome ( $0.75 \pm 0.05$ ;  $p < 0.0001$ ), clinical data ( $0.75 \pm 0.06$ ;  $p < 0.0001$ ), and DWMA data ( $0.76 \pm 0.05$ ;  $p < 0.0001$ ). This model also achieved the highest Pearson's correlation coefficient  $r$  of  $0.63 \pm 0.05$  ( $p < 0.0001$ ). This was significantly greater than using functional connectome data with a  $r$  of  $0.38 \pm 0.06$  ( $p < 0.0001$ ), structural connectome data with a  $r$  of  $0.45 \pm 0.07$  ( $p < 0.0001$ ), clinical data with a  $r$  of  $0.41 \pm 0.06$  ( $p < 0.0001$ ), and DWMA data with a  $r$  of  $0.38 \pm 0.05$  ( $p < 0.0001$ ). These support our hypothesis that integrating multimodal MRI and clinical data improves early prediction of motor deficits at 2 years corrected age in VPIs over using individual data modalities.

## Most Discriminative Feature Identification

Figure 2 shows the most discriminative region-to-region functional connections ranked by the proposed deep multimodal learning model for the prediction of cognitive, language, and motor deficits. Among 13 functional connections discriminative for at least two deficits, 8% are within the right hemisphere and 23% are within the left hemisphere only. Interhemispheric connections account for 69% of top discriminative connections. More detailed predictive functional connections to the individual deficits are shown in Supplementary Figures 1–3. Functional brain connections contributing to the prediction of all three

**TABLE 4 |** Performance comparison shows that our proposed deep multimodal learning model that uses combined feature sets (i.e., functional connectome + structural connectome + clinical data + DWMA) obtained at term-equivalent age outperforms each individual feature set for early identification of very preterm infants at high-risk for cognitive deficits and predicting their actual Bayley III Cognitive scores at 2 years corrected age.

Features	Cognitive deficit risk stratification				Cognitive Bayley III score prediction		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	<i>r</i>	MAE	SD of AE
Functional connectivity	72.0 ± 4.6	63.4 ± 7.7	80.5 ± 6.0	0.74 ± 0.05	0.34 ± 0.07	19.5	12.6
Structural connectivity	79.9 ± 4.1	73.5 ± 5.7	86.4 ± 6.0	0.81 ± 0.06	0.44 ± 0.05	13.3	11.8
Clinical features	77.3 ± 4.2	73.8 ± 5.1	82.3 ± 6.2	0.77 ± 0.05	0.41 ± 0.05	15.5	12.6
DWMA	80.6 ± 5.7	78.3 ± 4.8	82.8 ± 5.0	0.74 ± 0.05	0.39 ± 0.04	14.9	13.1
<b>Combined</b>	<b>88.4 ± 3.7</b>	<b>83.4 ± 6.5</b>	<b>90.3 ± 5.2</b>	<b>0.87 ± 0.05</b>	<b>0.62 ± 0.04</b>	<b>11.7</b>	<b>8.6</b>

**TABLE 5 |** Performance comparison shows that our proposed deep multimodal learning model using combined feature sets (i.e., functional connectome + structural connectome + clinical data + DWMA) obtained at term-equivalent age outperforms each individual feature set for early identification of very preterm infants at high-risk for language deficits and predicting their actual Bayley III Language scores at 2 years corrected age.

Features	Language deficit risk stratification				Language Bayley III score prediction		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	<i>r</i>	MAE	SD of AE
Functional Connectivity	74.8 ± 3.9	71.3 ± 5.4	78.2 ± 6.4	0.76 ± 0.04	0.44 ± 0.06	14.4	12.1
Structural Connectivity	76.2 ± 4.7	74.1 ± 7.2	78.2 ± 6.1	0.75 ± 0.05	0.48 ± 0.05	12.9	11.1
Clinical features	75.9 ± 4.2	74.2 ± 6.5	81.3 ± 4.2	0.78 ± 0.04	0.40 ± 0.06	13.7	10.4
DWMA	78.4 ± 4.2	75.3 ± 5.7	81.5 ± 5.9	0.78 ± 0.05	0.39 ± 0.06	15.4	10.2
<b>Combined</b>	<b>87.2 ± 5.3</b>	<b>83.6 ± 6.4</b>	<b>89.6 ± 6.1</b>	<b>0.85 ± 0.04</b>	<b>0.63 ± 0.04</b>	<b>10.5</b>	<b>8.2</b>

DWMA, diffuse white matter abnormality; *r*, Pearson's correlation coefficient; MAE, mean absolute error; SD of AE, standard deviation of absolute error.

**TABLE 6 |** Performance comparison shows that our proposed deep multimodal learning model using combined feature sets (i.e., functional connectome + structural connectome + clinical data + DWMA) obtained at term-equivalent age outperforms each individual feature set for early identification of very preterm infants at high-risk for motor deficits and predicting their actual Bayley III Motor scores at 2 years corrected age.

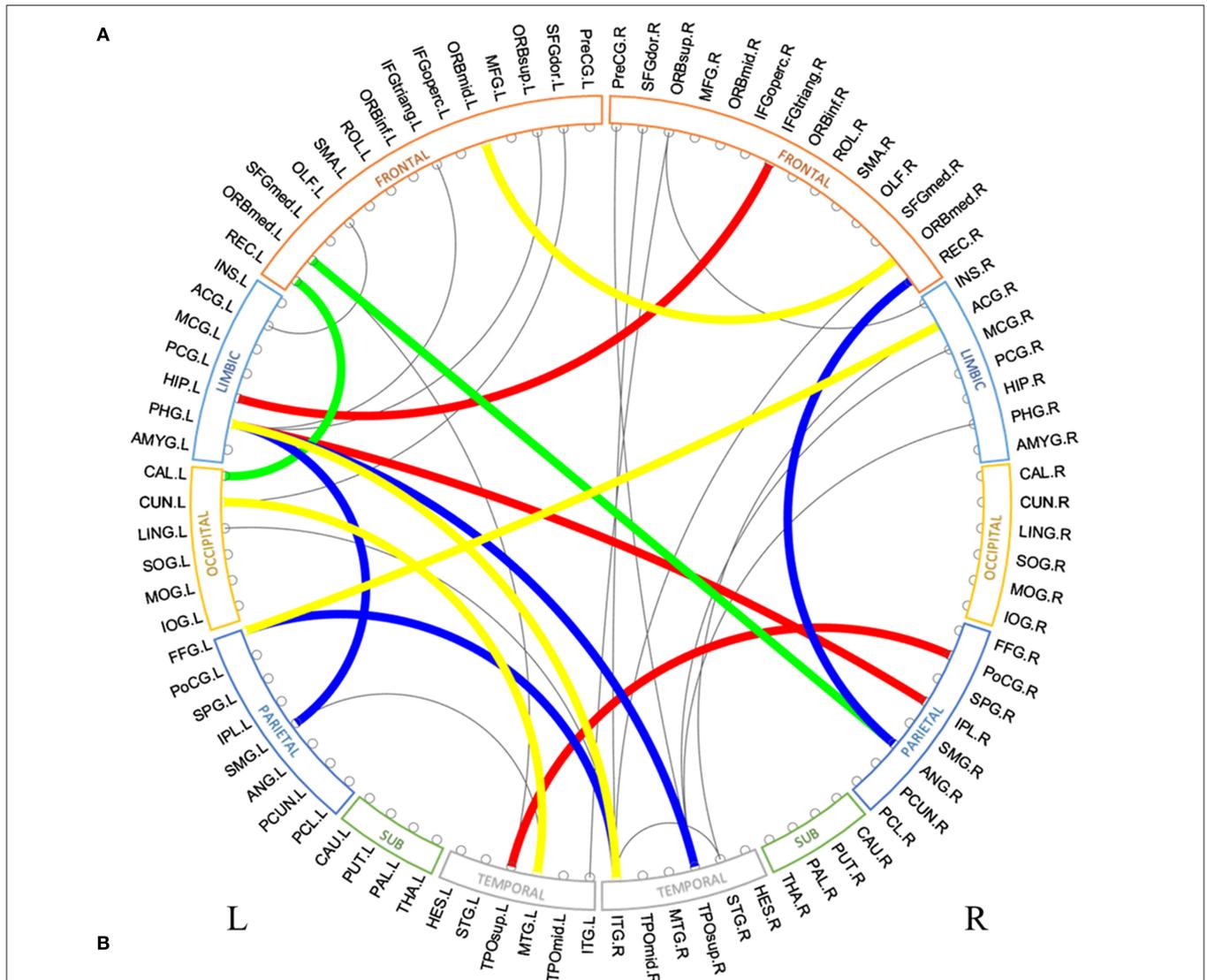
Features	Motor deficit risk stratification				Motor Bayley III score prediction		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	<i>r</i>	MAE	SD of AE
Functional connectivity	68.6 ± 4.8	70.9 ± 6.3	66.2 ± 8.1	0.71 ± 0.05	0.38 ± 0.06	15.9	12.5
Structural connectivity	74.1 ± 5.2	71.3 ± 8.2	77.0 ± 6.1	0.75 ± 0.06	0.45 ± 0.07	13.7	12.4
Clinical features	75.4 ± 5.3	73.4 ± 5.7	78.2 ± 6.3	0.75 ± 0.06	0.41 ± 0.06	13.8	10.7
DWMA	77.1 ± 4.7	80.7 ± 4.8	73.4 ± 6.4	0.76 ± 0.05	0.38 ± 0.05	14.5	11.1
<b>Combined</b>	<b>86.7 ± 5.2</b>	<b>87.6 ± 5.8</b>	<b>82.5 ± 4.9</b>	<b>0.85 ± 0.06</b>	<b>0.63 ± 0.05</b>	<b>11.6</b>	<b>9.2</b>

DWMA, diffuse white matter abnormality; *r*, Pearson's correlation coefficient; MAE, mean absolute error; SD of AE, standard deviation of absolute error.

deficits span frontal, limbic, occipital, temporal, and parietal lobes.

Similarly, **Figure 3** shows the most predictive *structural* connections ranked by the proposed deep multimodal learning model for the prediction of all three deficits. Among 13 structural connections discriminative for at least two deficits, 62% are within the right hemisphere and 23% are within the left hemisphere. Fifteen percent of top discriminative connections are interhemispheric connections. Structural brain connections contributing to the prediction of all three deficits focus on frontal, limbic, and parietal lobes, as well as subcortical gray nuclei. More detailed predictive structural connections to the individual deficits are shown in **Supplementary Figures 4–6**.

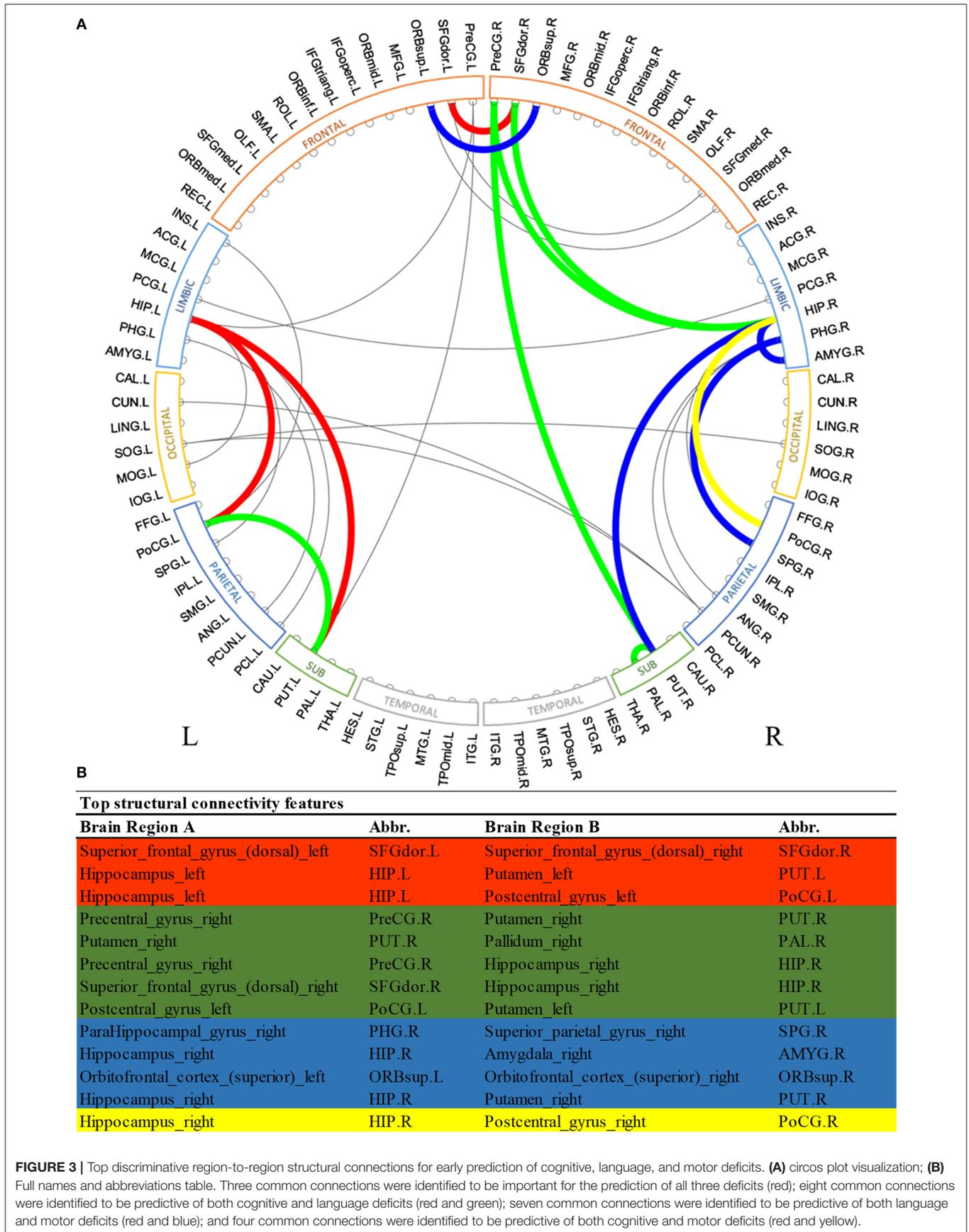
**Table 7** shows the discriminative clinical features ranked by our deep multimodal learning model for the prediction of all three neurodevelopmental (cognitive, language, and motor) deficits. As expected, several well-known neurodevelopment-relevant clinical features were repeatedly selected by the model as discriminative features for all three prediction tasks, such as mother's highest educational level, infant positive pressure respiratory therapy, head circumference at birth, birth weight, and gestational age at birth. Among 11 severity levels of DWMA feature, we found that threshold  $\alpha = 1.8$  DWMA feature was ranked as the most predictive DWMA feature for all three prediction tasks.



**Top functional connectivity features**

Brain Region A	Abbr.	Brain Region B	Abbr.
Postcentral_gyrus_right	PoCG.R	Temporal_pole_(superior)_left	TPOsup.L
ParaHippocampal_gyrus_left	PHG.L	Inferior_parietal_lobule_right	IPL.R
Inferior_frontal_gyrus_(triangular)_right	IFGtriang.R	Hippocampus_left	HIP.L
Rectus_gyrus_left	REC.L	Calcarine_cortex_left	CAL.L
Orbitofrontal_cortex_(medial)_left	ORBsupmed.L	Angular_gyrus_right	ANG.R
ParaHippocampal_gyrus_left	PHG.L	Supramarginal_gyrus_left	SMG.L
ParaHippocampal_gyrus_left	PHG.L	Temporal_pole_(superior)_right	TPOsup.R
Fusiform_gyrus_left	FFG.L	Inferior_temporal_gyrus_right	ITG.R
Rectus_gyrus_right	REC.R	Angular_gyrus_right	ANG.R
Anterior_cingulate_gyrus_right	ACG.R	Fusiform_gyrus_left	FFG.L
ParaHippocampal_gyrus_left	PHG.L	Inferior_temporal_gyrus_right	ITG.R
Orbitofrontal_cortex_(middle)_left	ORBmid.L	Orbitofrontal_cortex_(medial)_right	ORBsupmed.R
Cuneus_left	CUN.L	Middle_temporal_gyrus_left	MTG.L

**FIGURE 2 |** Top discriminative region-to-region functional connections for early prediction of cognitive, language, and motor deficits. **(A)** circo plot visualization; **(B)** Full names and abbreviations table. Three common connections were identified to be important for the prediction of all three deficits (red); five common connections were identified to be predictive of both cognitive and language deficits (red and green); seven common connections were identified to be predictive of both language and motor deficits (red and blue); and seven common connections were identified to be predictive of both cognitive and motor deficits (red and yellow).



**FIGURE 3 |** Top discriminative region-to-region structural connections for early prediction of cognitive, language, and motor deficits. **(A)** circos plot visualization; **(B)** Full names and abbreviations table. Three common connections were identified to be important for the prediction of all three deficits (red); eight common connections were identified to be predictive of both cognitive and language deficits (red and green); seven common connections were identified to be predictive of both language and motor deficits (red and blue); and four common connections were identified to be predictive of both cognitive and motor deficits (red and yellow).

**TABLE 7** | Top discriminative clinical features for early prediction of cognitive, language, and motor deficits.

Cognition		Language		Motor	
Features	Weights	Features	Weights	Features	Weights
Head circumference at birth (cm)	0.9	Birth weight (grams)	1.0	Mothers age	1.1
Mothers age	0.9	Head circumference at birth (cm)	0.8	Birth weight (grams)	1.0
Gestational age at birth (total weeks)	0.8	Gestational age at birth (weeks)	0.8	Parenteral alimentation (days)	0.9
Parenteral alimentation (days)	0.8	Highest education level	0.8	Maternal smoking status	0.9
Pulmonary hypertension history	0.8	Mothers age	0.8	PRBC transfusions	0.9
Birth weight (grams)	0.8	PRBC transfusions	0.7	Total time on oxygen therapy (days)	0.8
Fertility treatment	0.7	Birth length (cm)	0.7	Head circumference at birth (cm)	0.8
Birth length (cm)	0.7	Pulmonary hypertension history	0.7	Birth length (cm)	0.8
Gestational age at birth (weeks)	0.7	CS history before 35 weeks PMA	0.7	Patent ductus arteriosus history	0.8
Highest education level	0.7	Referral to child protective services	0.7	Pulmonary hypertension history	0.8
Pulmonary hemorrhage	0.6	Total time on oxygen therapy (days)	0.7	Ventilation therapy	0.8
Ventilation therapy	0.6	Maternal annual income	0.6	Respiratory support type at 36-week	0.8
Necrotizing enterocolitis	0.6	CS history at or after 35 weeks PMA	0.6	Pulmonary hemorrhage	0.7
Iron supplementation	0.6	Parenteral alimentation (days)	0.6	Gestational age at birth (weeks)	0.7
GI surgery that resulted in short gut	0.6	Retinopathy of prematurity exam	0.6	Iron supplementation	0.7
PRBC transfusions	0.6	Pulmonary hemorrhage	0.6	GI surgery that resulted in short gut	0.7
Total time on oxygen therapy (days)	0.6	Placental pathology	0.6	Pneumothorax	0.7
Hyperthyroidism	0.6	Birth hospital	0.6	Total number of days on CPAP	0.7
Surfactant	0.6	Total number of days on CPAP	0.5	Necrotizing enterocolitis	0.7
Seizure history	0.5	Chorioamnionitis	0.5	Infections history	0.7

## DISCUSSION

### Brain Connectome Data Are Predictive of Neurodevelopmental Deficits

There is an increasing consensus that human brain can be modeled as a complex network both at a structural as well as functional level (Stam et al., 2016). Structural networks typically represent connection pathways corresponding to white matter tracks between pairs of brain regions, measuring white matter integrity. Functional networks represent magnitudes of temporal cross-correlations between blood-oxygen-level dependent (BOLD) signals, measuring coupling strength. Neurodevelopmental deficits can be understood as dysconnectivity syndromes, therefore the quantifications of the abnormal structural and functional network using graph theory may enable neurodevelopmental prognosis. In VPIs, we have previously established correlations of later neurodevelopmental outcomes with at term obtained functional connectivity features derived from rs-fMRI (Gozdas et al., 2018); and structural connectivity features derived from DTI (Chen et al., 2020). In this work, our results showed both structural and functional connectivity features obtained at term-equivalent age are predictive of abnormal cognitive, language, and motor outcomes at 2 years corrected age. Our results also suggest that the predictive power of structural connectivity features is stronger than functional connectivity features. The significant performance improvement supports our hypothesis that integrating multimodal MRI and clinical data improves early prediction of cognitive, language, and motor deficits

independently, at 2 years corrected age in VPIs over using each individual data modality.

Recent advances in deep learning techniques, based on artificial neural networks (ANN), have made it possible to extract physiologically meaningful features and reveal new discriminative information from high dimensional MRI data (Hjelm et al., 2014; Plis et al., 2014; Mostapha and Styner, 2019). Applications of deep learning to analyze high-dimensional objectively-quantified connectome features derived from DTI, and rs-fMRI data may detect brain structural and functional abnormalities and tissue pathologies that are not readily visible to the human eye, thereby facilitating risk stratification (Kassner and Thornhill, 2010; Mostapha and Styner, 2019; Sahiner et al., 2019). There is a growing interest in developing deep learning approaches to predict a variety of brain disorders and neurodevelopmental deficits using MRI data (Wee et al., 2012; Kawahara et al., 2017; Gilmore et al., 2018; He et al., 2018; Heinsfeld et al., 2018; Girault et al., 2019; Saha et al., 2020). However, early prediction of neurodevelopmental deficits for preterm infants is a very challenging task. For example, Kawahara et al. (2017) developed a BrainNetCNN model to predict cognitive and motor developmental outcome scores from brain structural connectome with a Person's correlation coefficient  $r$  of 0.188 and 0.310, respectively. In another study, Saha et al. (2020) achieved a mean accuracy of 73% on predicting motor outcome in preterm infants by applying a CNN model on DTI data. Similarly, we previously developed a transfer learning neural network model using functional connectome data to predict cognitive outcome at 2 years of corrected age, achieved

an accuracy of 70.6% (He et al., 2018). These studies using single modality data demonstrated that deep learning models were promising tools, but there is still a long way ahead. In the current work, we demonstrated that deep multimodal learning model is able to significantly improve prediction performance by integrating multiple data modalities. This facilitates the early prediction of neurodevelopmental deficits for preterm infants in the clinical setting using deep learning models and multimodal data.

## Potential Brain Connectome Biomarkers at Birth of Later Neurodevelopment

We observed multiple common functional brain connections, bridging brain regions within bilateral frontal lobe, left limbic system, left temporal lobe, and right parietal lobe, that significantly contributed to the prediction of all three neurodevelopmental deficits at 2 years corrected age (**Figure 2**). These regions serve important functions for language, sensory, motor, and cognitive function. For example, our proposed model identified the functional connection between the right postcentral gyrus and superior part of left temporal pole in all prediction tasks. The postcentral gyrus is located within the parietal lobe and is adjacent to the precentral gyrus of the frontal lobe (which was also selected). It is the primary somatosensory cortex and the main sensory receptive area (Hyvärinen and Poranen, 1978). On the other hand, the temporal pole is involved in high level semantic representation and socio-emotional processing (Olson et al., 2007). It is conceivable that the network between these brain regions is involved in cognitive, language, and motor functions as assessed by the Bayley III standardized tests at 2 years corrected age. Several other regions that are well-established hubs for these three core functions, such as the inferior temporal gyrus, inferior frontal gyrus, and cuneus were also identified as predictive biomarkers by our multimodal model. These results highlight the self-taught learning capability of the proposed deep multimodal learning model.

In terms of structural brain connectome, we also found multiple common connections that significantly contributed to decision-making of all three neurodevelopmental deficits at 2 years corrected age (**Figure 3**). Bilateral putamen regions were associated with some of these discriminative structural connections. Putamen is a critical subcortical nuclei that regulates movement and learning (de Jong et al., 2008). Significant microstructural or macrostructural alterations of putamen have been associated with neurodevelopmental and neurodegenerative disorders, including developmental language impairment (Lee et al., 2013), Parkinson's disease (Menke et al., 2009), and epilepsy (Keller et al., 2011; Gerdes et al., 2012). For example, Keller et al. (2011) demonstrated increased fractional anisotropy and decreased volume of the putamen region in patients with juvenile myoclonic epilepsy. Furthermore, the fractional anisotropy of putamen was showed to be significantly correlated with age in prior studies (Snook et al., 2005; Silk et al., 2009). This enables putamen to be a potential biomarker of human brain developmental trajectory. In another study, Fischi-Gómez et al. (2015) showed that decreased connectivity

between basal ganglia (caudate, putamen, and globus pallidum combined) with frontal or parietal regions was associated with cognitive and emotional development in school age extremely preterm infants. We also previously demonstrated that the lenticular nucleus (combined putamen and globus pallidum) is ~15% smaller in extremely low birth weight infants as compared to full-term controls (Parikh et al., 2013). Apparently, our model took advantage of discriminative information embedded in the putamen-related structural connections for the neurodevelopmental prediction in this current work. Anatomically, the putamen is closely connected to the pallidum region. The short-range structural connection between putamen and pallidum within the right hemisphere was identified by our model to be predictive for both cognitive and language deficits. Our finding is consistent with several previous studies in non-VPI populations that highlighted the synchronization and dyssynchronization of putamen and pallidum (Cheruel et al., 1994; de Jong et al., 2008; Gooijers et al., 2016). Noteworthy, our model identified the structural connection between putamen and hippocampus within the left hemisphere for all neurodevelopmental deficits risk stratification, but only associated the mirror connection within the right hemisphere to language and motor deficits. It might be interesting to further investigate the mechanism behind such differences between structural connections linking putamen and hippocampus of left and right hemispheres.

The hippocampus was repeatedly identified by our models for all three prediction tasks using both brain functional and structural connectome data. The hippocampus is well-known for its primary role in organizing and storing information, and particularly in forming new memories (Kesner, 2007; Ekstrom and Ranganath, 2018). Prior studies reported that patients with mild Alzheimer's disease exhibited altered hippocampal activity on functional MRI during memory tasks (Small et al., 1999; Sperling, 2007). In a DTI study, mean diffusivity of the hippocampus was significantly associated with verbal memory performance (den Heijer et al., 2012). Our model appears to recognize the importance of hippocampus structurally and functionally. Our findings support the idea that the hippocampus plays a critical role in learning and cognition during early infancy (Beauchamp et al., 2008). These further indicate that our proposed deep multimodal learning model is capable of automatically learning and identifying neurologically meaningful functional and structural connectivity for prediction tasks of neurodevelopmental deficits. Intriguingly, the model identified multiple structural connections related to bilateral hippocampi, while it only recognized one functional connection associated with the hippocampus region within the left hemisphere of the brain. This may be due to fact that a multimodal integrative machine learning model tends to learn and utilize complementary features, instead of duplicated information. It is also notable that over half of the top discriminative functional connections were long-range connections across bilateral hemispheres, but only a small portion (15%) of structural connections were interhemispheric. Further investigation is needed to explore the influence of long-range functional connections

and short-range structural connections on neurodevelopment of neonates.

## Identified Clinical and DWMA Predictors of Later Neurodevelopment

We identified several antepartum, intrapartum, and postnatal clinical factors that were predictive of one or more neurodevelopmental outcome at 2 years corrected age. Most of these factors have been shown in one or more prior studies to be predictive of such outcomes, including gestational age, birth growth parameters, duration of oxygen therapy/respiratory support and cognitive, language, and motor outcomes (Ambalavanan et al., 2012; Linsell et al., 2015, 2016; Parikh et al., 2020). These predictors that are consistent with prior research demonstrate the self-taught learning capability of our deep multimodal learning model on discovering useful knowledge from high dimensional big data. For DWMA features, the threshold  $\alpha = 1.8$  DWMA feature was ranked as the most predictive feature for all three prediction tasks. In a prior independent study, we also found that threshold  $\alpha = 1.8$  DWMA feature is significantly correlated with 2 years cognitive and language outcomes (Parikh et al., 2020). Importantly, the proposed deep multimodal learning model ranked these clinical predictors by simultaneously considering functional and structural connectome features. Thus, the rankings of these predictors do not necessarily reflect their individual predictive power on neurodevelopment. In other words, the most predictive variable in a univariable analysis may not be ranked as the top discriminative feature by our models.

## Related to Deep Multimodal Learning

It has been long recognized that the integration of multimodal features improves the performance of machine learning methods. Each feature modality has its own characteristic, which is different from others, leading to the complexity of heterogeneous data. Therefore, the key factor in multimodal fusion task is how to fill the heterogeneity gap of different modalities. For example, in this work, the problem is how to fuse heterogeneous features (i.e., very high dimensional structural and functional connectome data, as well as, low dimensional clinical and DWMA data) in a multimodal setting. In other words, how one solves the challenge of fusing high-dimensional and low-dimensional data will significantly impact the final results (Xu et al., 2016). If integrated directly, low-dimensional data would be completely overwhelmed by high-dimensional data. Instead, we proposed to encode each unimodal data *via* an independent neural network. By varying the architecture of the individual neural network, we reduced the dimensions of the high-dimensional data, and augmented or maintained the dimensions of the low-dimensional data. We then projected the encoded representations with equal dimensions into a shared semantic subspace, where the multimodal features/representations can be aggregated into a single feature/representation vector. Such learned vector is expected to fuse complementary and supplementary semantics from different modalities. The advantages of the multimodal learning strategy we proposed include: (1) convenience of fusing several modalities and (2)

the shared common subspace tends to be modality-invariant, which is helpful for transferring knowledge from one modality to another (Guo et al., 2019).

## Study Limitations

This study has several limitations. First, though we have previously demonstrated that joint prediction of multiple neurodevelopmental deficits improves performance over independent prediction of each individual deficit (He et al., 2020), we opted to go with the latter approach in this work, since the training augmentation algorithm we used were not supported for multi-task label simulation. Second, the multimodal predictive feature identification was conducted based on the optimal multimodal neural network architecture rather than the optimal unimodal neural network. That is, the identified predictive unimodal features were constrained by the other modalities, therefore such feature identification schema cannot be used to infer the separated predictive features for each modality. Third, the current study is mainly about outcome prediction, more systematic statistical analysis will be needed to determine if brain connectome, DWMA or certain clinical risk factors are biomarkers for later neurodevelopment. Fourth, an atlas without the cerebellum was used for brain connectome quantification, however, functional and structural connections within the cerebellum may also be important for emerging functional outcomes. Finally, this current study should be considered as proof-of-concept due to the limited sample size. A larger population is necessary to test the model generalizability.

## CONCLUSION

We presented a novel deep multimodal learning framework integrating features derived from anatomical MRI, rs-fMRI, DTI, and clinical data obtained at term-equivalent age to predict Bayley-III developmental scores and identify very preterm infants at-high risk of developing cognitive, language, and motor deficits at 2 years corrected age. We demonstrated the value of multimodal MRI features as potential biomarkers for prediction of later neurodevelopmental deficits. We also reported a set of predictive functional and structural connections and clinical risk factors of neurodevelopmental deficits. A larger study with external validation is important to validate our approach to further assess its clinical utility and overall generalizability.

## DATA AVAILABILITY STATEMENT

Requests to access the data sets used in this study should be directed to the corresponding author with a formal data sharing agreement and approval from the requesting researcher's local ethics committee. Requests to access these datasets should be directed to Lili He (lili.he@cchmc.org).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Boards of the Nationwide Children's Hospital and Cincinnati Children's Hospital Medical

Center. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

LH: conceptualization, methodology, validation, formal analysis, visualization, writing—original draft, and funding acquisition. HL: methodology, software, validation, formal analysis, visualization, and writing—review and editing. MC: software, validation, visualization, and formal analysis. JW, MA, and JD: validation and writing—review and editing. NP: conceptualization, resources, validation, writing—review, editing, and funding acquisition. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Institutes of Health (R01-EB029944, R01-EB030582, R21-HD094085, R01-NS094200, and R01-NS096037). The funders played no role in the design, analysis, or presentation of the findings.

## ACKNOWLEDGMENTS

We sincerely thank our collaborators from the Cincinnati Infant Neurodevelopment Early Prediction Study (CINEPS)

## REFERENCES

- Ambalavanan, N., Carlo, W. A., Tyson, J. E., Langer, J. C., Walsh, M. C., Parikh, N. A., et al. (2012). Outcome trajectories in extremely preterm infants. *Pediatrics* 130, e115–e125. doi: 10.1542/peds.2011-3693
- Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M., and Grafton, S. T. (2011). Conserved and variable architecture of human white matter connectivity. *Neuroimage* 54, 1262–1279. doi: 10.1016/j.neuroimage.2010.09.006
- Bayley, N. (2009). *Bayley-III: Bayley Scales of Infant and Toddler Development*. San Antonio, TX: Giunti OS; Pearson.
- Beauchamp, M. H., Thompson, D. K., Howard, K., Doyle, L. W., Egan, G. F., Inder, T. E., et al. (2008). Preterm infant hippocampal volumes correlate with later working memory deficits. *Brain* 131, 2986–2994. doi: 10.1093/brain/awn227
- Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A. B., Narwal, R., et al. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* 379, 2162–2172. doi: 10.1016/S0140-6736(12)60820-4
- Boardman, J. P., Craven, C., Valappil, S., Counsell, S. J., Dyet, L. E., Rueckert, D., et al. (2010). A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. *Neuroimage* 52, 409–414. doi: 10.1016/j.neuroimage.2010.04.261
- Cao, M., Huang, H., and He, Y. (2017). Developmental connectomics from infancy through early childhood. *Trends Neurosci.* 40, 494–506. doi: 10.1016/j.tins.2017.06.003
- Chaddad, A., Desrosiers, C., and Toews, M. (2017). Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age. *Sci. Rep.* 7:45639. doi: 10.1038/srep45639
- Chen, M., Li, H., Wang, J., Yuan, W., Altaye, M., Parikh, N. A., et al. (2020). Early prediction of cognitive deficit in very preterm infants using brain structural connectome with transfer learning enhanced deep convolutional neural networks. *Front. Neurosci.* 14:858. doi: 10.3389/fnins.2020.00858

Investigators: Principal Investigator: NP, DO, MS. Collaborators (in alphabetical order): MA, PhD, Anita Arnsperger, RRT, Traci Beiersdorfer, RN BSN, Kaley Bridgewater, RT(MR) CNMT, Tanya Cahill, MD, Kim Cecil, PhD, Kent Dietrich, RT, Christen Distler, BSN RNC-NIC, Juanita Dudley, RN BSN, Brianne Georg, BS, Cathy Grisby, RN BSN CCRC, Lacey Haas, RT(MR) CNMT, Karen Harpster, PhD, OT/RL, LH, PhD, Scott K. Holland, PhD, V. S. Priyanka Illapani, MS, Kristin Kirker, CRC, Julia E. Kline, PhD, Beth M. Kline-Fath, HL, PhD, Matt Lanier, RT(MR) RT(R), Stephanie L. Merhar, MD MS, Greg Muthig, BS, Brenda B. Poindexter, MD MS, David Russell, JD, Kari Tepe, BSN RNC-NIC, Leanne Tamm, PhD, Julia Thompson, RN BSN, Jean A. Tkach, PhD, JW, PhD, Brynne Williams, RT(MR) CNMT, Kelsey Wineland, RT(MR) CNMT, Sandra Wuertz, RN BSN CCRP, Donna Wuest, AS, Weihong Yuan, PhD. We sincerely thank Jennifer Notestine, RN and Valerie Marburger, NNP for serving as our Nationwide Children's study coordinators and Mark Smith, MS, for serving as the study MR technologist. We are most grateful to the families that made this study possible.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2021.753033/full#supplementary-material>

- Cheruel, F., Dormont, J. F., Amalric, M., Schmiech, A., and Farin, D. (1994). The role of putamen and pallidum in motor initiation in the cat. I. Timing of movement-related single-unit activity. *Exp. Brain Res.* 100, 250–266. doi: 10.1007/BF00227195
- de Jong, L. W., van der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G., Bollen, E. L., et al. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* 131, 3277–3285. doi: 10.1093/brain/awn278
- den Heijer, T., der Lijn, F., Vernooij, M. W., de Groot, M., Koudstaal, P. J., van der Lugt, A., et al. (2012). Structural and diffusion MRI measures of the hippocampus and memory performance. *Neuroimage* 63, 1782–1789. doi: 10.1016/j.neuroimage.2012.08.067
- Ekstrom, A. D., and Ranganath, C. (2018). Space, time, and episodic memory: the hippocampus is all over the cognitive map. *Hippocampus* 28, 680–687. doi: 10.1002/hipo.22750
- Fischi-Gómez, E., Vasung, L., Meskaldji, D. E., Lazeyras, F., Borradori-Tolsa, C., Hagmann, P., et al. (2015). Structural brain connectivity in school-age preterm infants provides evidence for impaired networks relevant for higher order cognitive skills and social cognition. *Cereb. Cortex* 25, 2793–2805. doi: 10.1093/cercor/bhu073
- Fortin, J. P., Parker, D., Tunc, B., Watanabe, T., Elliott, M. A., Ruparel, K., et al. (2017). Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170. doi: 10.1016/j.neuroimage.2017.08.047
- Friston, K. J. (1994). *Statistical Parametric Mapping*. Cambridge, MA: Academic Press.
- Gerdes, J. S., Keller, S. S., Schwindt, W., Evers, S., Mohammadi, S., and Deppe, M. (2012). Progression of microstructural putamen alterations in a case of symptomatic recurrent seizures using diffusion tensor imaging. *Seizure* 21, 478–481. doi: 10.1016/j.seizure.2012.03.015
- Gilmore, J. H., Knickmeyer, R. C., and Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nat. Rev. Neurosci.* 19:123. doi: 10.1038/nrn.2018.1

- Girault, J. B., Munsell, B. C., Puechmaille, D., Goldman, B. D., Prieto, J. C., Styner, M., et al. (2019). White matter connectomes at birth accurately predict cognitive abilities at age 2. *Neuroimage* 192, 145–155. doi: 10.1016/j.neuroimage.2019.02.060
- Gooijers, J., Chalavi, S., Beeckmans, K., Michiels, K., Lafosse, C., Sunaert, S., et al. (2016). Subcortical volume loss in the thalamus, putamen, and pallidum, induced by traumatic brain injury, is associated with motor performance deficits. *Neurorehabil. Neural Repair* 30, 603–614. doi: 10.1177/1545968315613448
- Gozdas, E., Parikh, N. A., Merhar, S. L., Tkach, J. A., He, L., and Holland, S. K. (2018). Altered functional network connectivity in preterm infants: antecedents of cognitive and motor impairments? *Brain Struct. Funct.* 223, 3665–3680. doi: 10.1007/s00429-018-1707-0
- Guo, W., Wang, J., and Wang, S. (2019). Deep multimodal representation learning: a survey. *IEEE Access* 7, 63373–63394. doi: 10.1109/ACCESS.2019.2916887
- Hackman, D. A., and Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends Cogn. Sci.* 13, 65–73. doi: 10.1016/j.tics.2008.11.003
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G. (2017). Learning from class-imbalanced data: review of methods and applications. *Expert Syst. Appl.* 73, 220–239. doi: 10.1016/j.eswa.2016.12.035
- Hallquist, M. N., Hwang, K., and Luna, B. (2013). The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *Neuroimage* 82, 208–225. doi: 10.1016/j.neuroimage.2013.05.116
- He, L., Li, H., Holland, S., Yuan, W., Altabe, M., and Parikh, N. (2018). Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework. *Neuroimage Clin.* 18, 290–297. doi: 10.1016/j.nicl.2018.01.032
- He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J. R., et al. (2020). A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci. Rep.* 10:15072. doi: 10.1038/s41598-020-71914-x
- He, L., and Parikh, N. A. (2013). Atlas-guided quantification of white matter signal abnormalities on term-equivalent age MRI in very preterm infants: findings predict language and cognitive development at two years of age. *PLoS One* 8:e85475. doi: 10.1371/journal.pone.0085475
- He, L., and Parikh, N. A. (2015). Aberrant executive and frontoparietal functional connectivity in very preterm infants with diffuse white matter abnormalities. *Pediatr. Neurol.* 53, 330–337. doi: 10.1016/j.pediatrneurol.2015.05.001
- He, L., and Parikh, N. A. (2016). Brain functional network connectivity development in very preterm infants: the first six months. *Early Hum. Dev.* 98, 29–35. doi: 10.1016/j.earlhumdev.2016.06.002
- Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *Neuroimage Clin.* 17, 16–23. doi: 10.1016/j.nicl.2017.08.017
- Hjelm, R. D., Calhoun, V. D., Salakhutdinov, R., Allen, E. A., Adali, T., and Plis, S. M. (2014). Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks. *Neuroimage* 96, 245–260. doi: 10.1016/j.neuroimage.2014.03.048
- Hyvärinen, J., and Poranen, A. (1978). Receptive field integration and submodality convergence in the hand area of the post-central gyrus of the alert monkey. *J. Physiol.* 283, 539–556. doi: 10.1113/jphysiol.1978.sp012518
- Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167 [arXiv preprint]*.
- Jarjour, I. T. (2015). Neurodevelopmental outcome after extreme prematurity: a review of the literature. *Pediatr. Neurol.* 52, 143–152. doi: 10.1016/j.pediatrneurol.2014.10.027
- Johnston, M. V. (2009). Plasticity in the developing brain: implications for rehabilitation. *Dev. Disabil. Res. Rev.* 15, 94–101. doi: 10.1002/ddrr.64
- Kassner, A., and Thornhill, R. (2010). Texture analysis: a review of neurologic MR imaging applications. *Am. J. Neuroradiol.* 31, 809–816. doi: 10.3174/ajnr.A2061
- Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. *Neuroimage* 146, 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046
- Keller, S. S., Ahrens, T., Mohammadi, S., Möddel, G., Kugel, H., Ringelstein, E. B., et al. (2011). Microstructural and volumetric abnormalities of the putamen in juvenile myoclonic epilepsy. *Epilepsia* 52, 1715–1724. doi: 10.1111/j.1528-1167.2011.03117.x
- Kesner, R. P. (2007). Behavioral functions of the CA3 subregion of the hippocampus. *Learn. Mem.* 14, 771–781. doi: 10.1101/lm.688207
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv:1412.6980 [arXiv preprint]*.
- Kwon, S. H., Vasung, L., Ment, L. R., and Huppi, P. S. (2014). The role of neuroimaging in predicting neurodevelopmental outcomes of preterm neonates. *Clin. Perinatol.* 41, 257–283. doi: 10.1016/j.clp.2013.10.003
- Lee, J. C., Nopoulos, P. C., and Bruce Tomblin, J. (2013). Abnormal subcortical components of the corticostriatal system in young adults with DLI: a combined structural MRI and DTI study. *Neuropsychologia* 51, 2154–2161. doi: 10.1016/j.neuropsychologia.2013.07.011
- Li, H., Parikh, N. A., Wang, J., Merhar, S., Chen, M., Parikh, M., et al. (2019). Objective and automated detection of diffuse white matter abnormality in preterm infants using deep convolutional neural networks. *Front. Neurosci.* 13:610. doi: 10.3389/fnins.2019.00610
- Linsell, L., Malouf, R., Morris, J., Kurinczuk, J. J., and Marlow, N. (2015). Prognostic factors for poor cognitive development in children born very preterm or with very low birth weight: a systematic review. *JAMA Pediatr.* 169, 1162–1172. doi: 10.1001/jamapediatrics.2015.2175
- Linsell, L., Malouf, R., Morris, J., Kurinczuk, J. J., and Marlow, N. (2016). Prognostic factors for cerebral palsy and motor impairment in children born very preterm or very low birthweight: a systematic review. *Dev. Med. Child Neurol.* 58, 554–569. doi: 10.1111/dmcn.12972
- Liu, Y., Feng, X., and Zhou, Z. (2016). Multimodal video classification with stacked contractive autoencoders. *Signal Process.* 120, 761–766. doi: 10.1016/j.sigpro.2015.01.001
- Lowe, M. J., and Sorenson, J. A. (1997). Spatially filtering functional magnetic resonance imaging data. *Magn. Reson. Med.* 37, 723–729. doi: 10.1002/mrm.1910370514
- Martin, J. A., Hamilton, B. E., Osterman, M. J. K., and Driscoll, A. K. (2019). Births: final data for 2018. *Natl. Vital Stat. Rep.* 68, 1–47.
- May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends Cogn. Sci.* 15, 475–482. doi: 10.1016/j.tics.2011.08.002
- Menke, R. A., Scholz, J., Miller, K. L., Deoni, S., Jbabdi, S., Matthews, P. M., et al. (2009). MRI characteristics of the substantia nigra in Parkinson's disease: a combined quantitative T1 and DTI study. *Neuroimage* 47, 435–441. doi: 10.1016/j.neuroimage.2009.05.017
- Moeskops, P., Išgum, I., Keunen, K., Claessens, N. H. P., van Haastert, I. C., Groenendaal, F., et al. (2017). Prediction of cognitive and motor outcome of preterm infants based on automatic quantitative descriptors from neonatal MR brain images. *Sci. Rep.* 7:2163. doi: 10.1038/s41598-017-02307-w
- Mostapha, M., and Styner, M. (2019). Role of deep learning in infant brain MRI analysis. *Magn. Reson. Imaging* 64, 171–189. doi: 10.1016/j.mri.2019.06.009
- Nordhov, S. M., Ronning, J. A., Dahl, L. B., Ulvund, S. E., Tunby, J., and Kaarensen, P. I. (2010). Early intervention improves cognitive outcomes for preterm infants: randomized controlled trial. *Pediatrics* 126, e1088–e1094. doi: 10.1542/peds.2010-0778
- Olden, J. D., and Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Modell.* 154, 135–150. doi: 10.1016/S0304-3800(02)00064-9
- Olson, I. R., Plotzker, A., and Ezzyat, Y. (2007). The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731. doi: 10.1093/brain/awm052
- Parikh, N. A., He, L., Priyanka Illapani, V. S., Altabe, M., Folger, A. T., and Yeates, K. O. (2020). Objectively diagnosed diffuse white matter abnormality at term is an independent predictor of cognitive and language outcomes in infants born very preterm. *J. Pediatr.* 220, 56–63. doi: 10.1016/j.jpeds.2020.01.034
- Parikh, N. A., Lasky, R. E., Kennedy, K. A., McDavid, G., and Tyson, J. E. (2013). Perinatal factors and regional brain volume abnormalities at term in a cohort of extremely low birth weight infants. *PLoS One* 8:e62804. doi: 10.1371/journal.pone.0062804
- Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., et al. (2014). Deep learning for neuroimaging: a validation study. *Front. Neurosci.* 8:229. doi: 10.3389/fnins.2014.00229

- Pogribna, U., Burson, K., Lasky, R. E., Narayana, P. A., Evans, P. W., and Parikh, N. A. (2014). Role of diffusion tensor imaging as an independent predictor of cognitive and language development in extremely low-birth-weight infants. *AJNR Am. J. Neuroradiol.* 35, 790–796. doi: 10.3174/ajnr.A3725
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048
- Qi, J., and Peng, Y. (2018). “Cross-modal bidirectional translation via reinforcement learning,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)* (Stockholm), 2630–2636.
- Rasiwasia, N., Costa Pereira, J., Coviello, E., Doyle, G., Lanckriet, G. R., Levy, R., et al. (2010). “A new approach to cross-modal multimedia retrieval,” in *Proceedings of the 18th ACM International Conference on Multimedia* (Florence), 251–260.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016). Generative adversarial text to image synthesis. *arXiv:1605.05396 [arXiv preprint]*.
- Rogers, C. E., Smyser, T., Smyser, C. D., Shimony, J., Inder, T. E., and Neil, J. J. (2016). Regional white matter development in very preterm infants: perinatal predictors and early developmental outcomes. *Pediatr. Res.* 79:87. doi: 10.1038/pr.2015.172
- Saha, S., Pagnozzi, A., Bourgeat, P., George, J. M., Bradford, D., Colditz, P. B., et al. (2020). Predicting motor outcome in preterm infants from very early brain diffusion MRI using a deep learning convolutional neural network (CNN) model. *Neuroimage* 215:116807. doi: 10.1016/j.neuroimage.2020.116807
- Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., et al. (2019). Deep learning in medical imaging and radiation therapy. *Med. Phys.* 46, e1–e36. doi: 10.1002/mp.13264
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 618–626.
- Shi, F., Salzwedel, A. P., Lin, W., Gilmore, J. H., and Gao, W. (2017). Functional brain parcellations of the infant brain and the associated developmental trends. *Cereb. Cortex* 28, 1358–1368. doi: 10.1093/cercor/bhx062
- Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J. H., Lin, W., et al. (2011). Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One* 6:e18746. doi: 10.1371/journal.pone.0018746
- Silk, T. J., Vance, A., Rinehart, N., Bradshaw, J. L., and Cunnington, R. (2009). Structural development of the basal ganglia in attention deficit hyperactivity disorder: a diffusion tensor imaging study. *Psychiatry Res.* 172, 220–225. doi: 10.1016/j.psychres.2008.07.003
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556 [arXiv preprint]*.
- Small, S. A., Perera, G. M., DeLaPaz, R., Mayeux, R., and Stern, Y. (1999). Differential regional dysfunction of the hippocampal formation among elderly with memory decline and Alzheimer's disease. *Ann. Neurol.* 45, 466–472. doi: 10.1002/1531-8249(199904)45:4<466::AID-ANA8>3.0.CO;2-Q
- Snook, L., Paulson, L. A., Roy, D., Phillips, L., and Beaulieu, C. (2005). Diffusion tensor imaging of neurodevelopment in children and young adults. *Neuroimage* 26, 1164–1173. doi: 10.1016/j.neuroimage.2005.03.016
- Sperling, R. (2007). Functional MRI studies of associative encoding in normal aging, mild cognitive impairment, Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 1097, 146–155. doi: 10.1196/annals.1379.009
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Stam, C. J., van Straaten, E. C. W., Van Dellen, E., Tewarie, P., Gong, G., Hillebrand, A., et al. (2016). The relation between structural and functional connectivity patterns in complex brain networks. *Int. J. Psychophysiol.* 103, 149–160. doi: 10.1016/j.ijpsycho.2015.02.011
- Thompson, D. K., Chen, J., Beare, R., Adamson, C. L., Ellis, R., Ahmadzai, Z. M., et al. (2016). Structural connectivity relates to perinatal factors and functional impairment at 7 years in children born very preterm. *Neuroimage* 134, 328–337. doi: 10.1016/j.neuroimage.2016.03.070
- Thompson, D. K., Lee, K. J., Egan, G. F., Warfield, S. K., Doyle, L. W., Anderson, P. J., et al. (2014). Regional white matter microstructure in very preterm infants: predictors and 7 year outcomes. *Cortex* 52, 60–74. doi: 10.1016/j.cortex.2013.11.010
- Thompson, P. M., Cannon, T. D., Narr, K. L., Van Erp, T., Poutanen, V.-P., Huttunen, M., et al. (2001). Genetic influences on brain structure. *Nat. Neurosci.* 4:1253. doi: 10.1038/nn758
- Valizadeh, S. A., Liem, F., Mérillat, S., Hänggi, J., and Jäncke, L. (2018). Identification of individual subjects on the basis of their brain anatomical features. *Sci. Rep.* 8:5611. doi: 10.1038/s41598-018-23696-6
- Wang, R., Benner, T., Sorensen, A. G., and Wedeen, V. J. (2007). “Diffusion toolkit: a software package for diffusion imaging data processing and tractography,” *Proceedings of the International Society for Magnetic Resonance in Medicine* (Berlin).
- Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Brownhyke, J. N., Potter, G. G., et al. (2012). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59, 2045–2056. doi: 10.1016/j.neuroimage.2011.10.015
- Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186. doi: 10.1016/j.neuroimage.2008.10.055
- Wu, L., Wang, Y., and Shao, L. (2018). Cycle-consistent deep generative hashing for cross-modal retrieval. *IEEE Trans. Image Process.* 28, 1602–1612. doi: 10.1109/TIP.2018.2878970
- Wu, S., Bondugula, S., Luisier, F., Zhuang, X., and Natarajan, P. (2014). “Zero-shot event detection using multi-modal fusion of weakly supervised concepts,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH), 2665–2672.
- Xu, K., Ba, J., Kiros, R., Courville, A., Salakhutdinov, R., Zemel, R., et al. (2015). Show, attend and tell: Neural image caption generation with visual attention. *arXiv:1502.03044 [arXiv preprint]*.
- Xu, T., Zhang, H., Huang, X., Zhang, S., and Metaxas, D. N. (2016). “Multimodal deep learning for cervical dysplasia diagnosis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds. S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, W. Wells (Cham: Springer), 115–123.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 He, Li, Chen, Wang, Altaye, Dillman and Parikh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.