

## Article

# A Semi-Supervised Graph Convolutional Network for Early Prediction of Motor Abnormalities in Very Preterm Infants

Hailong Li <sup>1,2,3,4</sup> , Zhiyuan Li <sup>1,5</sup>, Kevin Du <sup>1</sup>, Yu Zhu <sup>1</sup>, Nehal A. Parikh <sup>3,6</sup> and Lili He <sup>1,2,3,4,5,\*</sup>

<sup>1</sup> Imaging Research Center, Department of Radiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

<sup>2</sup> Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA

<sup>3</sup> Neurodevelopmental Disorders Prevention Center, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

<sup>4</sup> Artificial Intelligence Imaging Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

<sup>5</sup> Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221, USA

<sup>6</sup> Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA

\* Correspondence: lili.he@cchmc.org

**Abstract:** Approximately 32–42% of very preterm infants develop minor motor abnormalities. Earlier diagnosis soon after birth is urgently needed because the first two years of life represent a critical window of opportunity for early neuroplasticity in infants. In this study, we developed a semi-supervised graph convolutional network (GCN) model that is able to simultaneously learn the neuroimaging features of subjects and consider the pairwise similarity between them. The semi-supervised GCN model also allows us to combine labeled data with additional unlabeled data to facilitate model training. We conducted our experiments on a multisite regional cohort of 224 preterm infants (119 labeled subjects and 105 unlabeled subjects) who were born at 32 weeks or earlier from the Cincinnati Infant Neurodevelopment Early Prediction Study. A weighted loss function was applied to mitigate the impact of an imbalanced positive:negative (~1:2) subject ratio in our cohort. With only labeled data, our GCN model achieved an accuracy of 66.4% and an AUC of 0.67 in the early prediction of motor abnormalities, outperforming prior supervised learning models. By taking advantage of additional unlabeled data, the GCN model had significantly better accuracy (68.0%,  $p = 0.016$ ) and a higher AUC (0.69,  $p = 0.029$ ). This pilot work suggests that the semi-supervised GCN model can be utilized to aid early prediction of neurodevelopmental deficits in preterm infants.

**Keywords:** preterm infant; machine learning; deep learning; graph convolutional network; diffusion MRI; T2-weighted MRI; motor abnormality; cerebral palsy; neurodevelopment; prognosis



**Citation:** Li, H.; Li, Z.; Du, K.; Zhu, Y.; Parikh, N.A.; He, L. A Semi-Supervised Graph Convolutional Network for Early Prediction of Motor Abnormalities in Very Preterm Infants. *Diagnostics* **2023**, *13*, 1508. <https://doi.org/10.3390/diagnostics13081508>

Academic Editor: Nikolaos Dikaos

Received: 28 February 2023

Revised: 5 April 2023

Accepted: 19 April 2023

Published: 21 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Preterm birth poses a high risk of developing motor abnormalities compared to infants who are term born [1,2]. Very preterm infants (<32 weeks gestational age) are especially at increased risk [3]. Globally, an estimated 32–42% of very preterm infants develop minor motor abnormalities [4,5]. Typical motor abnormalities, such as cerebral palsy and delayed motor development, and poor fine motor skills, are the result of abnormal brain development or brain injury during the fetal or neonatal period [6–9]. However, major motor impairments typically cannot be accurately diagnosed until 1–2 years of age, and minor motor impairments take even longer to diagnose [6]. Earlier diagnosis soon after birth is urgently needed, as the first 2 years of life represents a critical window of opportunity for early neuroplasticity in preterm infants [7]. Accurate and early prediction of motor abnormalities enables us to target early interventions to the highest-risk infants during periods of optimal neuroplasticity to improve their quality of life.

Diffusion tensor imaging (DTI) is an advanced MRI technique used to detect how water molecules travel along the white matter tracts in the brain [10,11]. It allows a more complex investigation of brain white matter microstructure and macrostructure [3,12,13]. In recent years, several studies have used DTI-derived features to predict motor abnormalities in preterm infants [14–18]. Despite the promising results of machine learning, especially deep learning, in the early prediction of motor abnormalities in neonates, the previous studies solely focused on analyzing the features of individual subjects in isolation, neglecting potential pairwise interactions between subjects.

A graph provides a natural framework for capturing pairwise relationships between individuals [19–21]. In a graph, individuals can be represented as nodes and pairwise relationships can be represented as edges (binary or weighted). The recent advent of graph convolutional networks (GCNs) [22,23] has led to advanced disease classification applications that simultaneously consider features of individual subjects and relationships between subjects. For example, Parisot et al. [24] constructed sparse population graphs for the diagnosis of autism spectrum disorder (ASD) and Alzheimer’s disease (AD). In a more recent work, Jiang et al. [25] developed a hierarchical GCN to further improve performance in ASD and AD classification tasks. These promising results inspired us to explore the capabilities of GCN models in the task of predicting motor abnormalities in very preterm infants.

For node classification tasks, GCN models are actually trained in a semi-supervised learning manner, enabling us to arbitrarily integrate more data from subjects regardless of the availability of their labels [22]. Prior early motor abnormality prediction models were all supervised learning models that used labeled training data only [14–17]. Semi-supervised learning methods can produce considerable improvement in classification accuracy by using additional unlabeled data together with labeled data [26,27]. Thus, we set out to explore whether a semi-supervised GCN model is able to improve prediction of motor abnormalities by leveraging the available unlabeled data.

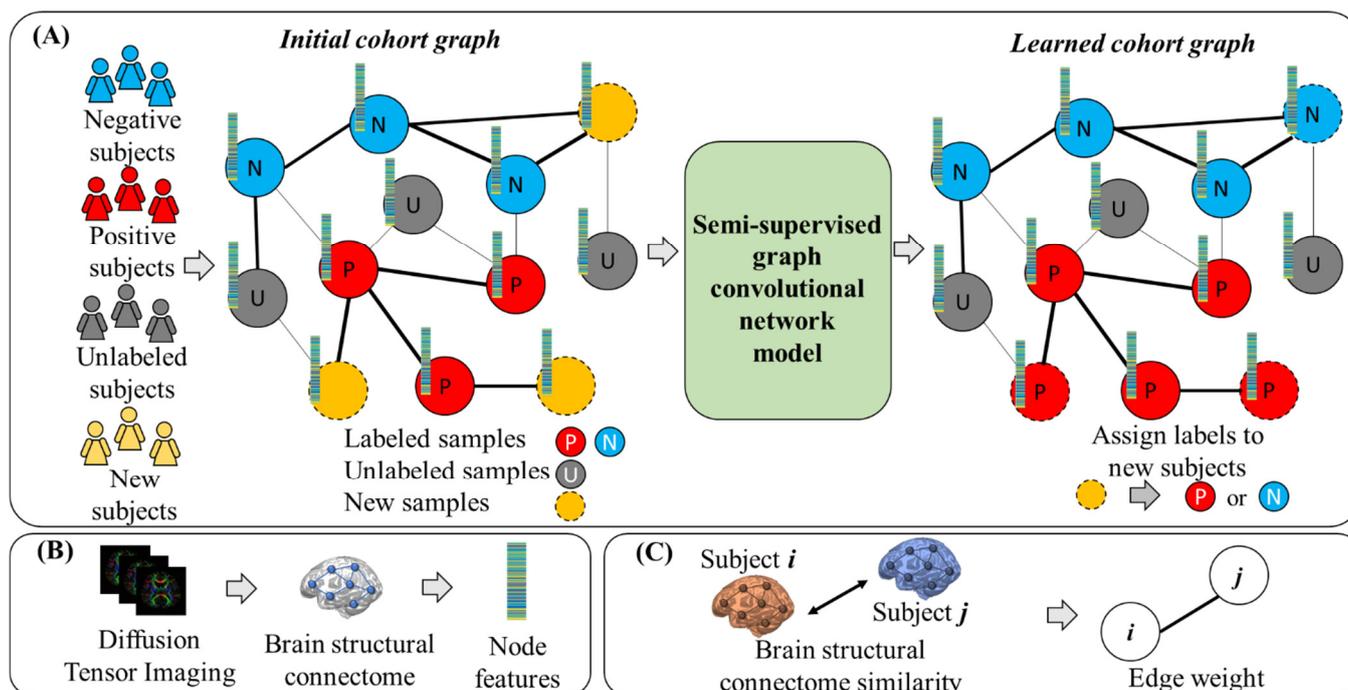
In this study, we developed a semi-supervised GCN model for predicting motor abnormalities at 2 years corrected age using DTI data obtained from very preterm infants at term-equivalent age. We conducted our work on a multisite regional cohort of preterm infants born at 32 weeks or earlier from the Cincinnati Infant Neurodevelopment Early Prediction Study (CINEPS). We sought to test our hypothesis of whether such a semi-supervised GCN model can improve prediction performance over peer supervised machine learning and deep learning models. We also hypothesized that the integration of labeled and unlabeled data will boost the performance of GCN models that used labeled DTI data alone.

## 2. Materials and Methods

### 2.1. Overview

An overview of our semi-supervised GCN framework for early risk prediction of motor impairment is illustrated in Figure 1. When this work commenced, only about half of the cohort had received 2-year motor assessment (labeled subjects), while the other half had not reached 2 years corrected age (unlabeled subjects). For subject (node) classification tasks, GCN models are trained in a semi-supervised learning manner, a strategy between supervised learning (with only labeled data) and unsupervised learning (with only unlabeled data). This allowed us to utilize the whole cohort regardless of their label availability. Given both labeled (blue and red) and unlabeled (gray) subjects ready for model learning, our main task was to classify new subjects (yellow) into one of the label groups. We first constructed an *initial cohort graph*, represented by a weighted graph  $\mathcal{G}(\square, \mathcal{E}, W)$  (Figure 1A). In this graph, each node  $\square$  represents a subject associated with a set of node features, and the weighted edge  $\mathcal{E}$  represents the similarity between two subjects associated with a weight  $W$  reflecting the similarity scale. This *initial cohort graph* includes labeled subjects (positive (P) and negative (N)), unlabeled subjects (U), and new to-be-classified subjects. We formulated our motor abnormalities prediction task

as a graph node binary classification problem, where the semi-supervised GCN model learned the *initial cohort graph* and assigned labels (i.e., positive for high risk and negative for low risk of developing motor abnormalities) to subjects in the *learned cohort graph*. We obtained DTI-derived brain structural connectomes and utilized vectorized structural connectivity as the node features (Figure 1B). We calculated the similarity between brain structural connectomes as the weight  $W(i, j)$  of edge  $\mathcal{E}$  between subjects  $i$  and  $j$  (Figure 1C). Additional details are elaborated in the cohort graph construction section.



**Figure 1.** The proposed semi-supervised learning framework to predict motor impairment at 2 years corrected age using brain structural connectomes derived from DTI data acquired at term in very preterm infants. Given both labeled (blue and red) and unlabeled (gray) subjects ready for model learning, our task was to classify new subjects (yellow) into one of the label groups. (A) We constructed an *initial cohort graph*, including labeled subjects (positive (P) and negative (N)), unlabeled (U) subjects, and to-be-predicted new subjects. The semi-supervised GCN model learned the *initial cohort graph* and assigned labels (i.e., positive for high risk and negative for low risk of developing motor impairment) to new subjects in the *learned cohort graph*. (B) We used diffusion tensor imaging (DTI)-derived brain structural connectomes as node features and (C) the similarity between brain connectomes as edge weights between nodes.

## 2.2. Data Acquisition and Processing

### 2.2.1. Subjects and MRI Data Acquisition

The Cincinnati Children’s Hospital Institutional Review Board approved this study. A parent or guardian of each infant gave written informed consent before enrollment. A total of 264 very preterm infants from five level-III Greater Cincinnati area neonatal intensive care units were enrolled at the time this work commenced. Cohort exclusion criteria included: (1) subjects with cyanotic heart disease or chromosomal or congenital anomalies affecting the central nervous system; (2) subjects who were hospitalized and mechanically ventilated on more than 50% supplemental oxygen at 45 weeks postmenstrual age (PMA). All experimental protocols involving human subjects were performed in accordance with the Declaration of Helsinki.

All study infants were imaged during unsedated sleep between 40 and 44 weeks postmenstrual age on a 3T Philips Ingenia scanner (Eindhoven, The Netherlands) with a 32-channel receiver head coil. All MRI scans were performed at Cincinnati Children’s

Hospital. A skilled neonatal nurse and neonatologist were both present for any scans requiring positive pressure airway support. The MRI data acquisition parameters were as follows: B800 DTI: echo time, 88 ms; repetition time, 6972 ms; flip angle,  $90^\circ$ ; field of view,  $160 \times 160 \text{ mm}^2$ ;  $80 \times 79$  matrix; 2 mm contiguous slices; scan time, 5:58 min. Thirty-six directions of diffusion gradients were applied with a b-value of  $800 \text{ s/mm}^2$ ; low b-value = 0 (4 b0 images were acquired with posterior–anterior phase encoding, and 1 b0 image was acquired with anterior–posterior phase encoding); axial T2-weighted MRI: echo time, 166 ms; repetition time, 18,567 ms; flip angle,  $90^\circ$ ; voxel dimensions,  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ ; scan time, 3:43 min.

### 2.2.2. MRI Data Preprocessing and Feature Extraction

We excluded 40 subjects due to severe brain injuries or large motion artifacts. The final cohort in this work contained 224 very preterm infants. For each subject, the DTI data were preprocessed using FMRIB's Diffusion Toolbox (in the FMRIB Software Library, FSL, FMRIB Analysis Group, 6.0.1, Oxford, UK). Diffusion tensor reconstruction and brain fiber tracking used Diffusion Toolkit (0.6.4, Massachusetts General Hospital, Boston, MA, USA)/TrackVis (0.6.1, Massachusetts General Hospital, Boston, MA, USA) [28]. Head motion and eddy current artifacts were mitigated by aligning all diffusion images to their B0 image via an affine transformation. The fiber tracking was performed employing a deterministic tracking algorithm and used an angular threshold of 35 degree with a fiber length threshold of 5 mm [29]. The whole-brain structural connectome was constructed based on 90 regions of interest (ROIs) defined from a neonatal automated anatomical labeling (AAL) atlas [30]. The structural connections between each pair of ROIs were calculated as the mean fractional anisotropy of each voxel intersecting the tract and then averaged over all tracts between the two ROIs, resulting in a  $90 \times 90$  symmetric adjacency matrix. This was performed using the UCLA Multimodal Connectivity Package (1.1, UCLA, Los Angeles, CA, USA) [31].

### 2.2.3. Motor Abnormalities Evaluation at 2 Years Corrected Age

Motor abnormalities evaluation was conducted for our subjects at 2 years corrected age as the gold-standard reference. One hundred and nineteen preterm infants had received the standardized Bayley Scales of Infant and Toddler Development III (Bayley-III) test at 2 years corrected age, and the remaining 105 infants had not yet reached this age at study commencement [32]. The Bayley-III Motor function composite (of fine and gross motor) sub-score was utilized to evaluate the risk of motor abnormalities for individual very preterm infants. The Bayley-III motor scores were normalized to a scale of 40–160, with a mean of 100 and a standard deviation of 15. We chose one standard deviation below the mean (i.e., 85) as the cutoff value to dichotomize the cohort into a high-risk group (motor score  $\leq 85$ ; 37 subjects) and a low-risk group (motor score  $> 85$ ; 82 subjects) in terms of developing moderate/severe motor abnormalities.

## 2.3. Semi-Supervised Graph Convolutional Networks

### 2.3.1. Construction of the Initial Cohort Graph

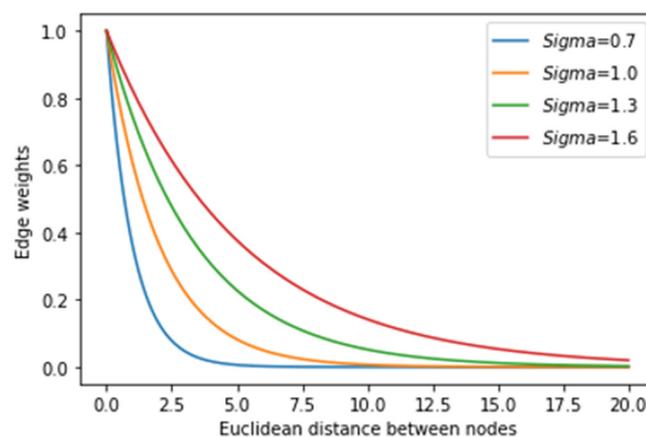
Graph construction is an important foundation for our semi-supervised learning framework. In this work, we constructed an *initial cohort graph* for the semi-supervised GCN to learn the features of individual subjects and capture their relationships. Specifically, we represented the very preterm infant cohort using a weighted graph  $\mathcal{G}(\mathcal{V}, \mathcal{E}, W)$ , where each node  $\mathcal{V}$  of the graph represented a very preterm infant, and the weighted edge  $\mathcal{E}$  represented the relationship between two infants associated with a weight  $W$ . This cohort graph contained all very preterm infants with or without 2-year motor function follow-up assessment tests. Each node  $\mathcal{V}$  was associated with a set of features. In our setting, we extracted and vectorized the unique connectivity weights of individual brain structural connectomes as a vector of 4005 features. Furthermore, we applied the natural exponential of Euclidean distance between node features (i.e., vectorized brain connectomes) as the

similarity/relationship between very preterm infants. As in a prior study [20], the edge weight between nodes  $i$  and  $j$  can be calculated by

$$W(i, j) = e^{-\frac{d(SC(i), SC(j))}{2\sigma^2}}$$

where  $d$  is the Euclidean distance,  $SC(i)$  and  $SC(j)$  are the brain structural connectome features of subjects  $i$  and  $j$ , and  $\sigma$  is a coefficient that determines the edge weight distribution.

The edge weight  $W$  is a number between 0 and 1. Figure 2 shows how edge weights between two nodes change according to the Euclidean distance between the brain connectomes of two nodes. When the Euclidean distance between node features is zero, that is, when two nodes have the same features, the edge weight is 1. When the Euclidean distance between node features increases, the edge weight decreases. When the Euclidean distance between node features is infinite, the edge weight becomes zero. The rationale is that subjects with the same labels (high-risk or low-risk motor impairments) tend to have more similar brain structural connectomes than subjects with different labels. The coefficient  $\sigma$  can be optimized to control the edge weight distribution for specific applications. In this work, we optimized graph construction using various coefficients  $\sigma = [0.7, 1.0, 1.3, 1.6, 1.9, 2.2]$  based on prediction performance. As such, we defined the *initial cohort graph* for our very preterm infants with explicit node features and edge weights.



**Figure 2.** Edge weights between two nodes are dependent on the Euclidean distance between the brain connectomes of two nodes. Various coefficients  $\sigma$  (i.e., sigma values) determine the edge weight distribution.

### 2.3.2. Architecture of the Graph Convolutional Network

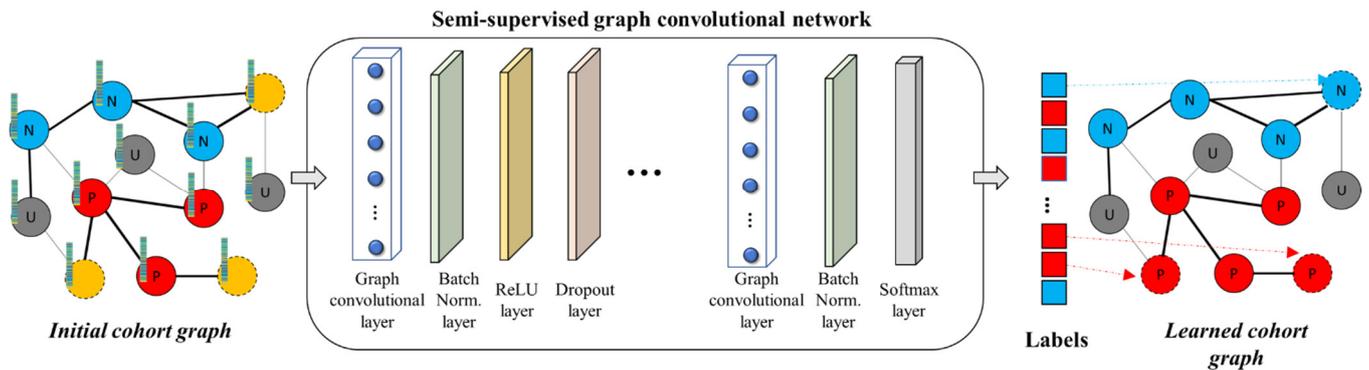
In this section, we elaborate the architecture of the semi-supervised GCN model for the early motor impairment risk prediction task. Given the constructed *initial cohort graph*  $\mathcal{G}(\mathcal{N}, \mathcal{E}, W)$ , the adjacency matrix of the graph is represented as  $A \in \mathbb{R}^{N \times N}$ , and the node feature matrix is represented as  $X \in \mathbb{R}^{N \times M}$ , where  $N$  is the number of subjects/nodes and  $M$  is the number of features (i.e., 4005 brain structural connectivity weights). The graph convolutional layer encodes the nodes with a forward propagation rule:

$$H^{l+1} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l$$

where  $\tilde{A} = A + I$  is the adjacency matrix with added self-connections and  $I$  is the identity matrix.  $W^l \in \mathbb{R}^{M \times F}$  is the weight matrix of graph convolutional filters in the layer  $l$ .  $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$  is the approximation of the normalized graph Laplacian of the weighted graph  $\mathcal{G}(\mathcal{N}, \mathcal{E}, W)$ , developed by [22] using a truncated expansion in terms of Chebyshev polynomials.  $H^{l+1}$  is the node feature embedding matrix of layer  $l + 1$ , aggregating the node feature embedding  $H^l$  from layer  $l$  using the graph adjacency matrix  $A$ , and  $H^0 = X$ .

We demonstrate the architecture of the developed semi-supervised GCN model in Figure 3. Our GCN model consists of a series of  $L$  graph learning blocks. Each of the

first  $L - 1$  graph learning blocks is made of a graph convolutional layer with  $k$  graph convolutional filters, a batch normalization layer, a rectified linear unit (ReLU) activation layer, and a dropout layer. At the end, we append a block with a graph convolutional layer, a batch normalization layer, and a SoftMax layer as the output of the GCN model for label classification. The number of hidden graph learning blocks is selected from [1, 2, 3, 4]. The number of graph convolutional filters is selected from [512, 256, 128, 64, 32, 16], according to a prior DTI study [33].



**Figure 3.** Architecture of the semi-supervised graph convolutional network (GCN) model. The model learns the whole cohort graph and outputs labels for individual nodes/subjects. Our GCN model consists of  $L$  hidden graph learning blocks and a final output block with a SoftMax layer for classifying graph nodes. Labeled samples are marked blue and red, Unlabeled samples are marked gray. New samples are marked yellow. P: positive, N: negative, U: unlabeled.

### 2.3.3. Model Training

The proposed semi-supervised GCN model can be trained using the binary cross-entropy loss function:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i|SC(i))) + (1 - y_i) \log(1 - p(y_i|SC(i)))$$

where  $p(y_i|SC(i))$  is the probability of the  $i$ th subject's structural connectome  $SC(i)$  being classified as the label  $y_i$ . However, since only a small portion of the cohort were at high risk for motor abnormalities, the sample ratio between the high-risk and low-risk motor impairment groups was highly imbalanced—a unique challenge that was not encountered in prior graph-based studies [24]. To mitigate the impact of such an imbalanced dataset issue, we considered a weighted binary cross-entropy loss function:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \beta_i y_i \log(p(y_i|SC(i))) + \beta_i (1 - y_i) \log(1 - p(y_i|SC(i)))$$

where  $p(y_i|SC(i))$  is the probability of the  $i$ th subject's structural connectome  $SC(i)$  being classified as the label  $y_i$ . The class weight  $\beta_i = \frac{N}{C_i}$  is computed using the inverse frequency method, where the variable  $C_i$  is the total number of samples for the class group of subject  $i$ .

To train the GCN model, we applied the Adam algorithm [34] to optimize the weights of the model with an initial learning rate of 0.01, the first and second decay rates being 0.9 and 0.999. We set a maximal epoch of 2000. During model training, the whole *initial cohort graph* (adjacency matrix  $A$ ) and their feature vectors (feature matrix  $X$ ) were inputted into the model. Only a training subset of the graph nodes was labeled. The remaining graph nodes (either unlabeled subjects or to-be-classified subjects in the testing set) can be seen by the GCN model without any label information during training. The features of these unlabeled nodes/subjects would impact the graph convolutions of labeled subjects when the GCN model performed the forward propagation. This makes the GCN model

learn both labeled subjects (features and labels) and unlabeled subjects (features only) in a semi-supervised manner.

#### 2.3.4. Model Evaluation

We evaluated the classification performance of the proposed semi-supervised GCN model through 5-fold cross-validation. The labeled subjects were partitioned into 5 portions. While one portion of labeled subjects was used for testing, the other 4 portions were used for model training. The unlabeled subjects (i.e., very preterm infants without 2-year follow-up tests) were included in the training procedure to aid the GCN model training and were not included in the testing procedure. This process was repeated 50 times to evaluate model performance variations.

We also compared our semi-supervised GCN model with several peer machine learning and deep learning models that have been applied in brain structural connectome studies [35,36], including a logistic regression model, a ridge classifier model, linear and nonlinear support vector machine (SVM) models, and deep neural networks. To handle imbalanced datasets, weighted variations of the aforementioned models were utilized if applicable. We trained weighted SVM models and also applied weighted cross-entropy loss functions to train the deep neural networks. After all of the models were trained, we evaluated the model performance by calculating the means and standard deviations (SDs) of five performance metrics, including accuracy, balanced accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC), for binary risk classification of motor abnormalities.

### 3. Results

#### 3.1. Subjects

We had a final cohort of 224 very preterm infants in this study after data exclusion (Section 2). Among them, 119 subjects had received the standardized Bayley III test at 2 years corrected age (labeled subjects), while the other 105 subjects (unlabeled subjects) had not reached 2 years of age when this study commenced. A cutoff value of 85 (i.e., one standard deviation below the mean Bayley III motor score) was utilized to dichotomize the labeled subjects into groups at high risk ( $\leq 85$ ; 37 subjects) and at low risk ( $>85$ ; 82 subjects) of developing moderate/severe motor abnormalities. Detailed demographics of the study cohort are listed in Table 1.

**Table 1.** Demographics of the subjects in the high-risk group, the low-risk group, and the unlabeled group.

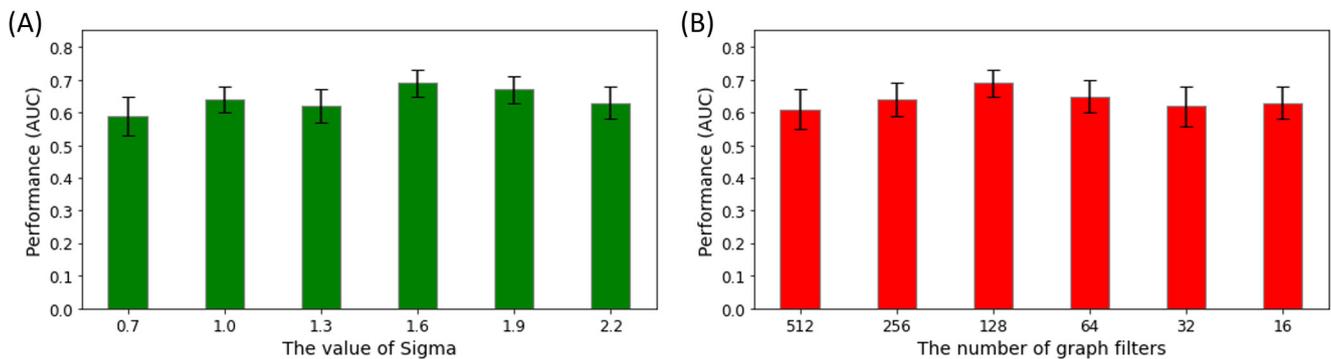
	High-Risk Group	Low-Risk Group	Unlabeled Group
Number of subjects	N = 37	N = 82	N = 105
Gestational age at birth (weeks)	28.9 (2.6)	29.3 (2.4)	29.5 (2.5)
Postmenstrual age at the scan (weeks)	42.7 (1.1)	42.0 (1.3)	43.0 (1.2)
Female, N (percentage)	11 (29.7%)	43 (52.4%)	45 (42.8%)
Birth weight (gram)	1244.7 (462.8)	1276.0 (424.4)	1353.7 (429.1)

Continuous variables are listed as mean (standard deviation).

#### 3.2. Semi-Supervised GCN Model Optimization

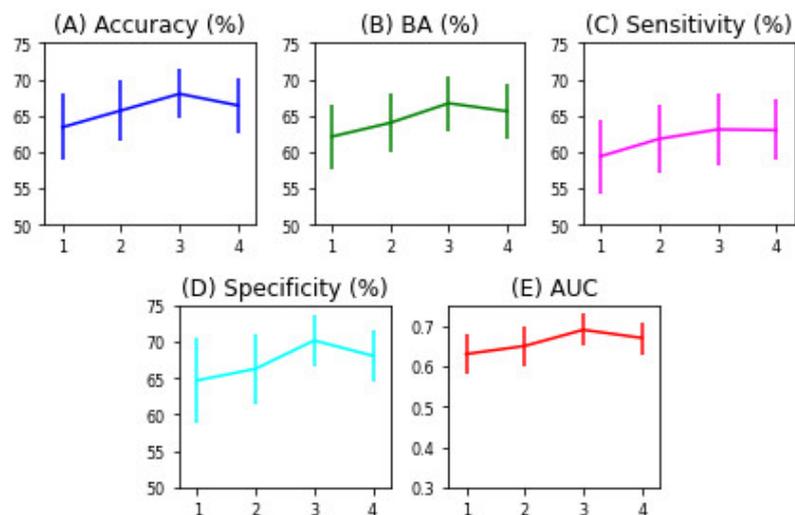
We first conducted optimization experiments to explore the best hyperparameters, including the weight coefficient  $\sigma$ , the number of graph filters, and the number of graph layers. Figure 4A demonstrates the model classification performance in terms of AUCs when the *initial cohort graph* was constructed using various weight coefficients  $\sigma$ . Given a GCN model with three layers and 128 graph filters in each layer, the best prediction AUC was achieved when the *initial cohort graph* was constructed using coefficient  $\sigma = 1.6$ . This is likely because the *initial cohort graph* ( $\sigma = 1.6$ ) was optimally constructed to model pairwise relationships between subjects. Similarly, Figure 4B shows the model's AUCs obtained by three-layer GCN models with different numbers of graph filters in individual layers. For the *initial cohort graph* ( $\sigma = 1.6$ ), the AUC of our GCN model increased as the number of

graph filters decreased. It reached peak performance ( $0.69 \pm 0.04$ ) when the number of filters was 128 and then decreased along with the decrease in the number of filters.



**Figure 4.** The model performance of the graph convolutional network model using (A) various edge weight coefficients for *initial cohort graph* construction and (B) different numbers of graph convolutional filters in individual layers.

Next, we continued to optimize the GCN model by searching for the optimal number of graph convolutional layers in the GCN model. We trained the GCN models with layers  $l = [1, 2, 3, 4]$ , and prediction performance with respect to motor impairment risk was evaluated on testing subjects using the *learned cohort graph*. As shown in Figure 5, the model performance increased as we increased the number of graph convolutional layers. The model performance started to drop when the number of layers in the GCN model reached four layers. Thus, we obtained an optimal architecture for our GCN model, which consists of three graph learning blocks. The first and second blocks are made of a graph convolutional layer with 128 graph convolutional filters, a batch normalization layer, a ReLU activation layer, and a dropout layer. The third block contains a graph convolutional layer with 128 graph filters, a batch normalization layer, and a SoftMax layer as the output of the model. The optimal accuracy ( $68.0\% \pm 3.4\%$ ), balanced accuracy ( $66.7\% \pm 3.8\%$ ), and AUC ( $0.69 \pm 0.04$ ) were achieved by this optimized GCN model.



**Figure 5.** The model performance of the graph convolutional network model using different numbers of graph convolutional layers. The x-axis represents the number of graph convolutional layers, while the y-axis indicates the model performance. (A) Accuracy. (B) Balanced accuracy (BA). (C) Sensitivity. (D) Specificity. (E) Area under the receiver operating characteristic curve (AUC).

### 3.3. Performance Comparison with Other Models

Next, we compared the proposed GCN model with several peer models (Section 2). Table 2 shows the model performance comparison for our early motor impairment prediction task. To investigate the effects of unlabeled data, we also trained the GCN model with and without unlabeled subjects. With the help of unlabeled data, the semi-supervised GCN model was able to identify subjects at high risk of developing moderate/severe motor impairment with a mean accuracy of 68.0% and a mean AUC of 0.69, outperforming the other peer models.

**Table 2.** Performance comparison for the prediction of very preterm infants at high risk vs. low risk of developing moderate/severe motor abnormalities at 2 years corrected age.

Models	Accuracy (%)	BA (%)	Sensitivity (%)	Specificity (%)	AUC
Logistic Regression	60.1 ± 4.0	58.7 ± 3.9	51.6 ± 4.4	65.7 ± 4.0	0.59 ± 0.03
Ridge Classifier	63.3 ± 3.7	61.2 ± 3.8	50.5 ± 4.2	71.9 ± 3.5	0.62 ± 0.04
SVM (linear kernel)	65.7 ± 3.4	64.5 ± 3.7	59.8 ± 5.1	69.1 ± 3.2	0.64 ± 0.03
SVM (rbf kernel)	66.2 ± 3.8	64.9 ± 4.0	61.5 ± 4.3	68.2 ± 4.1	0.65 ± 0.04
Deep neural network	65.9 ± 3.2	64.6 ± 4.4	60.2 ± 4.9	68.9 ± 4.3	0.64 ± 0.05
GCN (ours)	66.4 ± 3.3	65.4 ± 3.4	61.7 ± 4.7	69.0 ± 3.9	0.67 ± 0.05
GCN w/unlabeled (ours)	68.0 ± 3.4	66.7 ± 3.8	63.1 ± 4.9	70.2 ± 3.5	0.69 ± 0.04

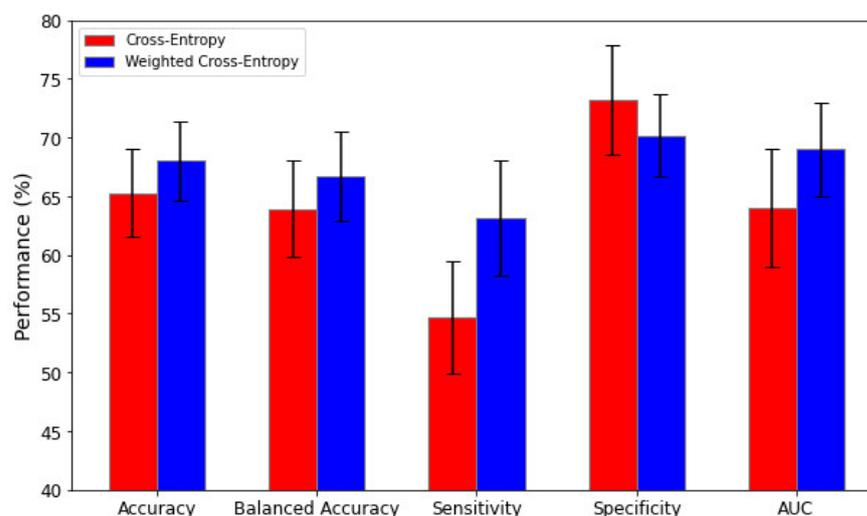
Data are presented as mean ± standard deviation. BA—Balanced Accuracy; AUC—area under the receive operating characteristic curve; SVM—Support vector machine; rbf—radial basis function; GCN—graph convolutional network. GCN w/unlabeled—GCN trained with labeled and unlabeled data. All other models were trained with labeled data in a supervised manner.

Among peer supervised learning models, the nonlinear SVM model achieved the best performance with a mean accuracy of 66.2% and an AUC of 0.65. Although the ridge classifier reached a specificity of 71.9%, this superior performance was only due to its low capability to handle the imbalanced dataset, and its sensitivity was undesirably low at 50.5%. Our semi-supervised GCN model without unlabeled data achieved a significantly higher AUC ( $p = 0.029$ ) than the nonlinear SVM model, although the accuracy difference was not significant ( $p = 0.476$ ). This validated our hypothesis that the semi-supervised GCN model is able to improve prediction performance over peer supervised machine learning and deep learning models.

Our GCN model with additional unlabeled data achieved improved prediction performance over the GCN model without using any unlabeled data. By taking advantage of 105 unlabeled data, the model had significantly better prediction performance in terms of accuracy (68.0% vs. 66.4%,  $p = 0.016$ ) and AUCs (0.69 vs. 0.67,  $p = 0.029$ ). Our semi-supervised GCN model also achieved better performance than a nonlinear SVM model, with an increase of 1.8% in accuracy ( $p = 0.014$ ) and 0.04 in AUC ( $p < 0.001$ ). This validated our hypothesis that the integration of labeled and unlabeled data is able to boost the performance of GCN models that used labeled data alone.

### 3.4. Impact of Weighted Loss Functions

To investigate the impact of weighted loss functions, we compared the performance of our semi-supervised GCN model using either typical cross-entropy or weighted cross-entropy loss functions. Figure 6 displayed the bar plots of five performance metrics from the GCN models with two different loss functions. The model using the typical cross-entropy loss function achieved a low sensitivity of  $54.7 \pm 4.8\%$  and a high specificity of  $73.2 \pm 4.7\%$ , indicating that this model tends to place more subjects in the low-risk group than in the high-risk group. In contrast, the GCN model using the weighted loss function reached a better equilibrium with an improved sensitivity of  $63.1 \pm 4.9\%$  and a slightly lower specificity of  $70.2 \pm 3.5\%$ . Such a better equilibrium was also reflected in the higher accuracy, balanced accuracy, and AUC of the semi-supervised GCN model trained with the weighted cross-entropy loss function.



**Figure 6.** The model performance comparison of the graph convolutional network model using cross-entropy and weighted cross-entropy loss functions.

#### 4. Discussion

Early identification of very preterm infants who may develop moderate/severe motor impairment is critical to design early personalized clinical intervention for improving their quality of life. In this work, we developed a semi-supervised GCN model for the early prediction of motor abnormalities at 2 years of age in very preterm infants using both labeled and unlabeled neuroimaging data. Our GCN model achieved a mean accuracy of 68.0% and a mean AUC of 0.69 on a partially labeled cohort of 224 very preterm infants and significantly outperformed multiple prior machine learning and deep learning models.

##### 4.1. Comparison of the GCN Model with Other Models

Several prior studies have used DTI-derived features to predict motor abnormalities of preterm infants using supervised learning models. Chau et al. [14] documented that multiple brain regional fractional anisotropy values measured by DTI were associated with adverse motor development of very preterm infants. Brown et al. [15] calculated topological features (e.g., node degrees, clustering coefficients, and global efficiencies) of brain structural connectomes derived from DTI data for individual very preterm infants, then applied SVM models to discriminate between normal and abnormally low scores of motor development. In another study, Kawahara et al. [16] proposed a deep learning framework with customized convolutional filters to predict motor developmental scores by using brain structural connectomes derived from DTI data for preterm infants. More recently, we developed a deep convolutional neural network (CNN)-based multi-modal learning model to integrate DTI data together with other data modalities to predict motor abnormalities in very preterm infants [17]. However, these prior supervised learning models [14–17] in the context of neurodevelopmental outcome prediction did not consider modeling pairwise interactions between subjects and rather relied on subject-specific imaging feature vectors, limiting the full potential of the discriminative ability of neuroimaging data.

Graph-based approaches have been successfully applied in disease diagnosis applications [20,21]. Zhao et al. [20] constructed a compact population graph by representing AD and normal control subjects as nodes and deriving edge weights using a local reconstruction method. Then, they conducted a node classification in the constructed population graph to classify AD patients. Later, Tong et al. [21] developed a nonlinear graph fusion approach to merge multi-modality data into a unified graph for AD classification. However, these earlier studies mainly focused on pairwise relationships between subjects and did not consider features of individual subjects. Compared to other models, the GCN model is able to simultaneously learn neuroimaging features of subjects and consider the pairwise similarity between neuroimaging features during the model training procedure. Thus,

GCN models have achieved state-of-the-art classification results in multiple applications for disease diagnosis. For example, Parisot et al. [24] applied GCN models in the diagnosis of ASD and AD. In this study, a vector of neuroimaging features was associated with nodes that represented subjects, and non-imaging features were used to model pairwise relationships between subjects as edge weights. Then, GCN models were applied to learn constructed population graphs and classify individual nodes into normal control or disease groups. Theoretically, GCNs could be considered as generalizations of the classic CNN model. Different from CNN models that operate on regular grid-like data/images using a fixed receptive field, GCN models are able to capture both local and global information in an irregular graph. Our GCN model outperformed other models in the early prediction of motor impairment in preterm infants. This is consistent with prior studies [20,21] using GCN models in other disease diagnosis applications.

#### *4.2. Architecture Optimization of the GCN*

In our work, we optimized our model by examining multiple hyperparameters, including the edge weight coefficient, the number of graph filters, and the number of graph layers. Our results showed that prediction performance reached a peak when our model had three graph convolutional layers. The performance started to decrease once we added more layers to the GCN model. The final GCN model contains three graph convolutional layers, each of which has 128 graph filters. The optimal depth of our GCN model depends on several factors, such as the complexity of the motor impairment prediction task. GCN models with one or two layers are easy to train with a smaller dataset due to fewer trainable weights. However, they may not have enough capacity to learn in such a complex prediction task. On the other hand, GCN models with more layers may have dramatically increased numbers of trainable weights of graph filters. Without a large dataset, simply adding more layers to the GCN model may cause an overfitting issue in our early-prediction task. Therefore, our GCN model with three layers reached the right balance between GCN model complexity and our dataset size.

#### *4.3. Semi-Supervised Learning Using Unlabeled Datasets*

Furthermore, most prior early-prediction studies mainly relied on labeled data to train supervised learning models. Typically, a large number of labeled training data are essential to develop robust models with desirable generalizability. Unfortunately, labeled training data are commonly very limited in most existing motor abnormalities prediction studies, since it is very challenging to collect large-scale neuroimaging datasets, especially ones with outcome data [37,38]. Acquiring later outcome data requires long-term follow-up of subjects to obtain clinical diagnoses. In this sense, a semi-supervised learning strategy that combines additional unlabeled data with labeled data during training may facilitate the development of prediction models. It is relatively easy to acquire unlabeled data, since clinical diagnoses or long-term follow-ups are not necessary. This means that it is feasible to significantly enrich a dataset for model development. Our results (Table 2) suggest that a semi-supervised GCN model is capable of leveraging additional unlabeled data to further improve the prediction of motor abnormalities.

#### *4.4. Weighted Loss Functions for Imbalanced Datasets*

The imbalanced dataset issue is a unique challenge in early-prediction studies of neurodevelopment in very preterm infants [17,35,39]. There are often significantly fewer subjects who develop moderate–severe motor abnormalities than who do not. Although this fact is desirable for infants, it creates a challenge for those developing machine learning models. When there are significantly more subjects in one group (referred to as the majority group), a model tends to assign undesirably more subjects to this majority group (i.e., low-risk motor impairment in this study) than to the minority group that contains fewer subjects (i.e., high-risk motor impairment). To reduce the impact of such imbalanced dataset issues, we applied a weighted cross-entropy loss function to train our semi-supervised

GCN model. The results showed that the weighted loss function was effective to handle this problem and prevent the model being biased toward the majority group.

#### 4.5. Limitations

There are certain limitations to this current study. First, the main limitation of the current work is that the semi-supervised GCN model represents a type of transductive learning, which requires training data to be available when the trained model infers an unseen new sample. This characteristic may limit the model's applications on certain low-memory devices, such as mobile devices. Second, the graph construction strategy is key for a GCN model to perform a node classification task. An inappropriately constructed graph may negatively impact the learning procedures of semi-supervised GCN models. To generalize the method and achieve optimal performance in other applications, tailored adjustments or modifications to the proposed graph construction may be necessary. Finally, although we demonstrated that our semi-supervised GCN model is able to mitigate the impact of small datasets, it is possible to further improve the model's performance by using additional labeled DTI data.

## 5. Conclusions

In this study, we strived to develop a semi-supervised GCN model for the early prediction of motor impairment at 2 years corrected age in preterm infants using both labeled and unlabeled DTI-derived brain structural connectomes. We demonstrated that the prediction performance of our proposed semi-supervised GCN model exceeded that of several prior supervised learning models. This semi-supervised GCN model can be generalized to aid early prediction of various neurodevelopmental deficits in very preterm infants.

**Author Contributions:** H.L.: Conceptualization, Methodology, Software, Validation, Data Analysis, Visualization, Writing—Original Draft. Z.L.: Formal Analysis, Writing—Review and Editing, K.D.: Formal Analysis, Writing—Review and Editing, Y.Z.: Formal Analysis, Writing—Review and Editing. N.A.P.: Conceptualization, Resources, Validation, Writing—Review and Editing. L.H.: Conceptualization, Methodology, Validation, Writing—Review and Editing, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Institutes of Health (R01-EB029944, R01-EB030582, R01-NS094200, and R01-NS096037) and Academic and Research Committee (ARC) Awards of Cincinnati Children's Hospital Medical Center. The funders played no role in the design, analysis, or presentation of the findings.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Cincinnati Children's Hospital Medical Center (2016-4483; date of approval: 26 July 2016).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Requests to access the datasets 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.

**Acknowledgments:** We sincerely thank our collaborators from the Cincinnati Infant Neurodevelopment Early Prediction Study (CINEPS) Investigators.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rogers, E.E.; Hintz, S.R. Early neurodevelopmental outcomes of extremely preterm infants. *Semin. Perinatol.* **2016**, *40*, 497–509. [[CrossRef](#)]
2. Hintz, S.R.; Barnes, P.D.; Bulas, D.; Slovis, T.L.; Finer, N.N.; Wrage, L.A.; Das, A.; Tyson, J.E.; Stevenson, D.K.; Carlo, W.A.; et al. Neuroimaging and Neurodevelopmental Outcome in Extremely Preterm Infants. *Pediatrics* **2015**, *135*, e32–e42. [[CrossRef](#)] [[PubMed](#)]
3. Parikh, N.A. Advanced neuroimaging and its role in predicting neurodevelopmental outcomes in very preterm infants. *Semin. Perinatol.* **2016**, *40*, 530–541. [[CrossRef](#)]

4. Williams, J.; Lee, K.J.; Anderson, P.J. Prevalence of motor-skill impairment in preterm children who do not develop cerebral palsy: A systematic review. *Dev. Med. Child Neurol.* **2010**, *52*, 232–237. [[CrossRef](#)]
5. Van Hus, J.W.; Potharst, E.S.; Jeukens-Visser, M.; Kok, J.H.; Van Wassenaer-Leemhuis, A.G. Motor impairment in very preterm-born children: Links with other developmental deficits at 5 years of age. *Dev. Med. Child Neurol.* **2014**, *56*, 587–594. [[CrossRef](#)] [[PubMed](#)]
6. McIntyre, S.; Morgan, C.; Walker, K.; Novak, I. Cerebral Palsy—Don't Delay. *Dev. Disabil. Res. Rev.* **2011**, *17*, 114–129. [[CrossRef](#)] [[PubMed](#)]
7. Parikh, N.A.; Harpster, K.; He, L.; Illapani, V.S.P.; Khalid, F.C.; Klebanoff, M.A.; O'shea, T.M.; Altaye, M. Novel diffuse white matter abnormality biomarker at term-equivalent age enhances prediction of long-term motor development in very preterm children. *Sci. Rep.* **2020**, *10*, 15920. [[CrossRef](#)]
8. Kline, J.E.; Illapani, V.S.P.; He, L.; Parikh, N.A. Automated brain morphometric biomarkers from MRI at term predict motor development in very preterm infants. *NeuroImage Clin.* **2020**, *28*, 102475. [[CrossRef](#)]
9. Chandwani, R.; Kline, J.E.; Harpster, K.; Tkach, J.; Parikh, N.A. The Cincinnati Infant Neurodevelopment Early Prediction Study (CINEPS) Group Early micro- and macrostructure of sensorimotor tracts and development of cerebral palsy in high risk infants. *Hum. Brain Mapp.* **2021**, *42*, 4708–4721. [[CrossRef](#)]
10. Alexander, A.L.; Lee, J.E.; Lazar, M.; Field, A.S. Diffusion tensor imaging of the brain. *Neurotherapeutics* **2007**, *4*, 316–329. [[CrossRef](#)] [[PubMed](#)]
11. O'donnell, L.J.; Westin, C.-F. An Introduction to Diffusion Tensor Image Analysis. *Neurosurg. Clin. N. Am.* **2011**, *22*, 185–196. [[CrossRef](#)] [[PubMed](#)]
12. Meoded, A.; Huisman, T. Diffusion Tensor Imaging of Brain Malformations: Exploring the Internal Architecture. *Neuroimaging Clin. N. Am.* **2019**, *29*, 423–434. [[CrossRef](#)] [[PubMed](#)]
13. Kline, J.E.; Dudley, J.; Illapani, V.S.P.; Li, H.; Kline-Fath, B.; Tkach, J.; He, L.; Yuan, W.; Parikh, N.A. Diffuse excessive high signal intensity in the preterm brain on advanced MRI represents widespread neuropathology. *Neuroimage* **2022**, *264*, 119727. [[CrossRef](#)] [[PubMed](#)]
14. Chau, V.; Synnes, A.; Grunau, R.E.; Poskitt, K.J.; Brant, R.; Miller, S.P. Abnormal brain maturation in preterm neonates associated with adverse developmental outcomes. *Neurology* **2013**, *81*, 2082–2089. [[CrossRef](#)]
15. Brown, C.J.; Miller, S.P.; Booth, B.G.; Poskitt, K.J.; Chau, V.; Synnes, A.R.; Zwicker, J.G.; Grunau, R.E.; Hamarneh, G. Prediction of motor function in very preterm infants using connectome features and local synthetic instances. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015.
16. Kawahara, J.; Brown, C.J.; Miller, S.P.; Booth, B.G.; Chau, V.; Grunau, R.E.; Zwicker, J.G.; Hamarneh, G. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *Neuroimage* **2017**, *146*, 1038–1049. [[CrossRef](#)]
17. He, L.; Li, H.; Chen, M.; Wang, J.; Altaye, M.; Dillman, J.R.; Parikh, N.A. Deep Multimodal Learning From MRI and Clinical Data for Early Prediction of Neurodevelopmental Deficits in Very Preterm Infants. *Front. Neurosci.* **2021**, *15*, 753033. [[CrossRef](#)]
18. Parikh, M.; Chen, M.; Braimah, A.; Kline, J.; McNally, K.; Logan, J.; Tamm, L.; Yeates, K.; Yuan, W.; He, L.; et al. Diffusion MRI Microstructural Abnormalities at Term-Equivalent Age Are Associated with Neurodevelopmental Outcomes at 3 Years of Age in Very Preterm Infants. *AJNR Am. J. Neuroradiol.* **2021**, *42*, 1535–1542. [[CrossRef](#)]
19. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994.
20. Zhao, M.; Chan, R.H.M.; Chow, T.W.S.; Tang, P. Compact Graph based Semi-Supervised Learning for Medical Diagnosis in Alzheimer's Disease. *IEEE Signal Process Lett.* **2014**, *21*, 1192–1196. [[CrossRef](#)]
21. Tong, T.; Gray, K.; Gao, Q.; Chen, L.; Rueckert, D. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognit.* **2017**, *63*, 171–181. [[CrossRef](#)]
22. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint* **2016**, arXiv:1609.02907.
23. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* **2016**, *29*. [[CrossRef](#)]
24. Parisot, S.; Ktena, S.I.; Ferrante, E.; Lee, M.; Guerrero, R.; Glocker, B.; Rueckert, D. Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Med. Image Anal.* **2018**, *48*, 117–130. [[CrossRef](#)] [[PubMed](#)]
25. Jiang, H.; Cao, P.; Xu, M.; Yang, J.; Zaiane, O. Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Comput. Biol. Med.* **2020**, *127*, 104096. [[CrossRef](#)]
26. Zhu, X.J. *Semi-Supervised Learning Literature Survey*; University of Wisconsin-Madison, Department of Computer Sciences: Madison, WI, USA, 2005.
27. Michaeli, T.; Eldar, Y.C.; Sapiro, G. Semi-supervised multi-domain regression with distinct training sets. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012.
28. Hess, C.P.; Mukherjee, P.; Han, E.T.; Xu, D.; Vigneron, D.B. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magn. Reson. Med.* **2006**, *56*, 104–117. [[CrossRef](#)] [[PubMed](#)]
29. Wang, R.; Benner, T.; Sorensen, A.G.; Wedeen, V.J. Diffusion toolkit: A software package for diffusion imaging data processing and tractography. *Proc. Intl. Soc. Mag. Reson Med.* **2007**, *15*, 3720.

30. Shi, F.; Yap, P.-T.; Wu, G.; Jia, H.; Gilmore, J.H.; Lin, W.; Shen, D. Infant Brain Atlases from Neonates to 1- and 2-Year-Olds. *PLoS ONE* **2011**, *6*, e18746. [[CrossRef](#)]
31. Bassett, D.S.; Brown, J.A.; Deshpande, V.; Carlson, J.M.; Grafton, S.T. Conserved and variable architecture of human white matter connectivity. *Neuroimage* **2011**, *54*, 1262–1279. [[CrossRef](#)]
32. Bayley, N. *Bayley-III: Bayley Scales of Infant and Toddler Development*; Giunti OS: Florence, Italy, 2009.
33. Zhang, X.; He, L.; Chen, K.; Luo, Y.; Zhou, J.; Wang, F. Multi-View Graph Convolutional Network and Its Applications on Neuroimage Analysis for Parkinson’s Disease. *AMIA Annu. Symp. Proc.* **2018**, *2018*, 1147–1156. [[PubMed](#)]
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Chen, M.; Li, H.; Wang, J.; Yuan, W.; Altaye, M.; Parikh, N.A.; He, L. Early Prediction of Cognitive Deficit in Very Preterm Infants Using Brain Structural Connectome with Transfer Learning Enhanced Deep Convolutional Neural Networks. *Front. Neurosci.* **2020**, *14*, 858. [[CrossRef](#)]
36. He, L.; Li, H.; Wang, J.; Chen, M.; Gozdas, E.; Dillman, J.R.; Parikh, N.A. A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci. Rep.* **2020**, *10*, 15072. [[CrossRef](#)] [[PubMed](#)]
37. He, L.; Li, H.; Holland, S.K.; Yuan, W.; Altaye, M.; Parikh, N.A. Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework. *NeuroImage Clin.* **2018**, *18*, 290–297. [[CrossRef](#)] [[PubMed](#)]
38. Li, H.; Parikh, N.A.; He, L. A Novel Transfer Learning Approach to Enhance Deep Neural Network Classification of Brain Functional Connectomes. *Front. Neurosci.* **2018**, *12*, 491. [[CrossRef](#)] [[PubMed](#)]
39. Ali, R.; Li, H.; Dillman, J.R.; Altaye, M.; Wang, H.; Parikh, N.A.; He, L. A self-training deep neural network for early prediction of cognitive deficits in very preterm infants using brain functional connectome data. *Pediatr. Radiol.* **2022**, *52*, 2227–2240. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.