Disentangled Hybrid Transformer for Identification of Infants with Prenatal Drug Exposure

Jiale Cheng[1,2], Zhengwang Wu[1], Xinrui Yuan[1], Li Wang[1], Weili Lin[1], Karen Grewen[3], and Gang Li[1(✉)]

[1] Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[2] Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University, Chapel Hill, NC 27599, USA
[3] Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
gang_li@med.unc.edu

Abstract. Prenatal drug exposure, which occurs during a time of extraordinary and critical brain development, is typically associated with cognitive, behavioral, and physiological deficits during infancy, childhood, and adolescence. Early identifying infants with prenatal drug exposures and associated biomarkers using neuroimages can help inform earlier, more effective, and personalized interventions to greatly improve later cognitive outcomes. To this end, we propose a novel deep learning model called disentangled hybrid volume-surface transformer for identifying individual infants with prenatal drug exposures. Specifically, we design two distinct branches, a volumetric network for learning non-cortical features in 3D image space, and a surface network for learning features on the highly convoluted cortical surface manifold. To better capture long-range dependency and generate highly discriminative representations, image and surface transformers are respectively employed for the volume and surface branches. Then, a disentanglement strategy is further proposed to separate the representations from two branches into complementary variables and common variables, thus removing redundant information and boosting expressive capability. After that, the disentangled representations are concatenated to a classifier to determine if there is an existence of prenatal drug exposures. We have validated our method on 210 infant MRI scans and demonstrated its superior performance, compared to ablated models and state-of-the-art methods.

Keywords: Prenatal Drug Exposure, Cortical Surface, Transformer.

1 Introduction

Prenatal drug exposure is a significant public concern and occurs during a time of extremely dynamic and critical brain development. It can lead to long-term cognitive and behavioral disadvantages that may persist throughout an individual's life and potentially into the next generation, underscoring the importance of early prognostication of

infants who are at risk for poor developmental outcomes [1]. Noninvasive MR imaging holds great potential in early identifying infants with prenatal drug exposure and revealing brain structural abnormalities and biomarkers associated with prenatal drug exposure. This will help inform earlier, more effective, and personalized interventions to greatly improve later cognitive outcomes in this highly vulnerable population.

However, this is very challenging because the intrinsic patterns for identifying an individual with prenatal drug exposure from the normal ones are overwhelmed by the rapid and complex brain development. Conventional methods [2,3] designed for neuroimage-based brain disorder diagnosis for adults and older children typically learn image-based features and only suit the scenario with subtle longitudinal brain changes and thus typically fail to work on dynamic infant brains. To tackle this issue, previous studies [4-6] take advantages of cortical surface-based features [7] instead of image-based features to capture the dynamic and complex neurobiological changes of the cerebral cortex during infancy, e.g., for infant cognition prediction [4]. As prenatal drug exposure can affect both cortical and deep noncortical regions, a flexible framework that can leverage cortical surfaces to capture complex, subtle cortical abnormal developmental patterns and MRI volumes to identify deficits in noncortical regions is critically desired.

Therefore, in this paper, we propose a novel deep learning method called disentangled hybrid volume-surface transformer (DHT) and apply it to identify infants with prenatal drug exposure, by taking advantages of both cortical surface-based representation and volumetric image-based representation. To this end, we transform them into an embedding space, where the regularity and variability of infant brain representations of surfaces and volumes can be effectively measured. Specifically, the Vision Transformer [8] and Spherical Surface Transformer [9] are chosen as the basic models to boost the discrimination capabilities for the volumed-based data and surface-based data, respectively, by leveraging their superior capabilities in modeling the long-range dependency. The motivation is that the Vision Transformer is well suitable for learning image-related features, especially for non-cortical regions, while the Spherical Surface Transformer is ideally capable of learning complex features on the cortical surface manifold with an intrinsic spherical topology. Then, we minimize the redundancy between surface-based and volume-based information and extract their complementary information to boost identification accuracy by separating their shared information from their specific information. To achieve this, we disentangle the latent variables of two encoders into representation-shared codes and representation-specific codes and enforce the representation-shared codes obtained from different representations to be as similar as possible, while the representation-specific codes to be different from each other as much as possible. This approach thus helps to not only offer a general, unified, and comparable embedding space by unifying hybrid representations in a single embedding space, but also effectively extract discriminative information for classification. To validate its effectiveness, we evaluated our method on 210 infant MRI scans to identify infants with prenatal opioid exposure and demonstrated its superior performance, compared to ablated models and state-of-the-art methods.

2 Method

2.1 Overview

As the schematic diagram shown in **Fig. 1**, our DHT works on the T1w and T2w MR images and cortical surfaces of both hemispheres (each with 40,962 vertices). The architecture has two main parts, including a hybrid Transformer-based encoding branch to produce effective feature representation for volume data and surface data in a unified embedding space, and a disentanglement block for semantically separating the redundant volume-surface shared information and the representative volume-specific and surface-specific information, finally boosting the identification accuracy.

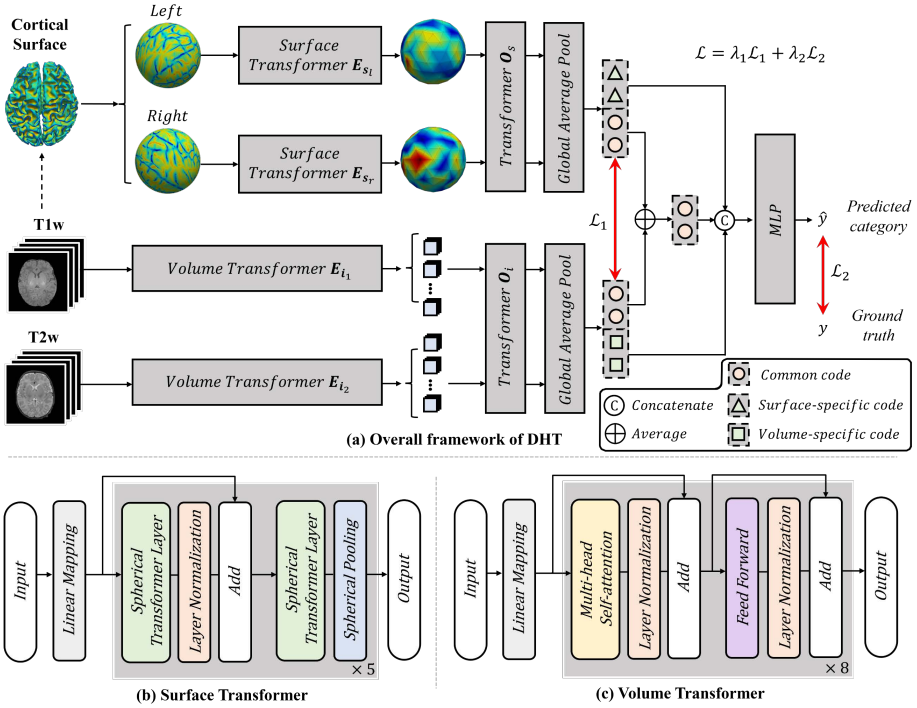

Fig. 1. The schematic diagram of our DHT that learns to embed the hybrid volume-surface data into a unified embedded space in an end-to-end architecture and further leverages the complementary information to boost the discriminative representation extraction.

Our data can be formulated as $(\boldsymbol{s}_l, \boldsymbol{s}_r, \boldsymbol{i}_1, \boldsymbol{i}_2, y)$, where $\boldsymbol{s}_l$ and $\boldsymbol{s}_r$ are the cortical surface feature maps for left hemisphere and right hemisphere, respectively, while $\boldsymbol{i}_1$ and $\boldsymbol{i}_2$ are T1w image and T2w image, respectively; y is the category of the existence of prenatal drug exposure.

2.2 Hybrid Volume-Surface Transformer

For each component in the input, we employ a neural network as its respective encoder $\boldsymbol{E_x}$, where $\boldsymbol{x} \in \{\boldsymbol{s_l}, \boldsymbol{s_r}, \boldsymbol{i_1}, \boldsymbol{i_2}\}$. The encoding branches $\boldsymbol{E_{s_l}}$ and $\boldsymbol{E_{s_r}}$ for cortical surfaces are five spherical transformer blocks as illustrated in **Fig. 1(b)**. Each spherical transformer block adopts a spherical transformer layer [9], which includes a 2-ring hexagonal multi-head self-attention layer, followed by layer normalization [10] and ReLU activation to extract vertex-wise representation, which are then downsampled by another spherical transformer layer and a hexagonal mean pooling layer [11,12] (except the last spherical transformer block) to serve as the input of the subsequent layer. On the other side, we design the encoding branches for $\boldsymbol{E_{i_1}}$ and $\boldsymbol{E_{i_2}}$ similarly as shown in **Fig. 1(c)**, each of which includes eight vision transformer blocks. Each block consists of a multi-head self-attention layer and a feed forward layer followed by the layer normalization and ReLU.

Based on the outputs from these four encoders, we propose a learnable spatial attention mechanism for localization of discriminative brain regions. Specifically, taken the surface branches as an example, let $\boldsymbol{f_{s_l}} = \boldsymbol{E_{s_l}}(\boldsymbol{s_l})$ and $\boldsymbol{f_{s_r}} = \boldsymbol{E_{s_r}}(\boldsymbol{s_r}) \in \mathbb{R}^{162\times C}$ be the vertex-wise representations (produced by the last spherical transformer block) for the left and right hemispheres, respectively. We first concatenate them as a matrix with the shape of $324 \times C$, where C is the dimension of the unified embedding space we created. Then, a self-attention operation [13] $\boldsymbol{O_s}$ is applied to capture cross-hemisphere long-range dependencies, which refines the vertex-wise representations from both hemispheric surfaces, resulting in a unified feature matrix $\boldsymbol{O_s}(\boldsymbol{f_{s_l}}, \boldsymbol{f_{s_r}}) \in \mathbb{R}^{324\times C}$. As shown in **Fig. 1(a)**, $\boldsymbol{O_s}(\boldsymbol{f_{s_l}}, \boldsymbol{f_{s_r}})$ is further input into a global average pool (GAP) layer to be a holistic feature vector $\boldsymbol{h_s} = GAP(\boldsymbol{O_s}(\boldsymbol{f_{s_l}}, \boldsymbol{f_{s_r}})) \in \mathbb{R}^C$ representing the latent variable of the whole cerebral cortex. Similarly, we can obtain the latent variable for the volume-based data as $\boldsymbol{h_i} = GAP(\boldsymbol{O_i}(\boldsymbol{f_{i_1}}, \boldsymbol{f_{i_2}}))$. Notably, by using $\boldsymbol{h_s}$ and $\boldsymbol{h_i}$ to identify infants with prenatal drug exposure, $\boldsymbol{O_s}$ and $\boldsymbol{O_i}$ highlight discriminative cortex regions on the surface and deep brain regions in the volume, respectively.

2.3 Latent Variable Disentanglement

To better learn the combined information from surfaces and volumes, shared and complementary information should be separated. Here, $\boldsymbol{h_m}$, where $m \in \{s, i\}$, is disentangled into two parts: $\boldsymbol{h_m^{Spe}}$ and $\boldsymbol{h_m^{Com}} \in \mathbb{R}^{C/2}$. $\boldsymbol{h_m^{Com}}$ is the common code representing the shared information amongst surfaces and volumes, while $\boldsymbol{h_m^{Spe}}$ is the specific code representing the complementary information that differentiates one from the other. The basic requirements of the disentanglement are: 1) the concatenation of $\boldsymbol{h_m^{Com}}$ and $\boldsymbol{h_m^{Spe}}$ equals $\boldsymbol{h_m}$; 2) $\boldsymbol{h_i^{Com}}$ and $\boldsymbol{h_s^{Com}}$ should be as similar as possible; 3) $\boldsymbol{h_i^{Spe}}$ should differ from $\boldsymbol{h_s^{Spe}}$ as much as possible. Therefore, $\mathcal{L}_1$ is defined as:

$$\mathcal{L}_1 = \mathcal{L}_{Com}/\mathcal{L}_{Spe}, \tag{1}$$

$$\mathcal{L}_{Com} = ||\boldsymbol{h_s^{Com}} - \boldsymbol{h_i^{Com}}||_2, \tag{2}$$

$$\mathcal{L}_{Spe} = ||\boldsymbol{h_s^{Spe}} - \boldsymbol{h_i^{Spe}}||_2. \tag{3}$$

Since each latent variable has been disentangled into the common code $\boldsymbol{h}_m^{Com}$ and the specific code $\boldsymbol{h}_m^{Spe}$, the combined information is formed as $\boldsymbol{h}_{s,i} = (\boldsymbol{h}_i^{Spe}, \boldsymbol{h}_{s,i}^{Com}, \boldsymbol{h}_s^{Spe})$, where $\boldsymbol{h}_{s,i}^{Com} = (\boldsymbol{h}_s^{Com} + \boldsymbol{h}_i^{Com})/2$. A multi-layer perceptron neural network (MLP) is then designed as a classifier to predict the category of each subject from $\boldsymbol{h}_{s,i}$. Finally, the objective function to end-to-end optimize DHT is written as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2, \tag{4}$$

$$\mathcal{L}_2 = -log\big(P(y|\boldsymbol{s}_l, \boldsymbol{s}_r, \boldsymbol{i}_1, \boldsymbol{i}_2; \theta)\big), \tag{5}$$

where $P(y|\boldsymbol{s}_l, \boldsymbol{s}_r, \boldsymbol{i}_1, \boldsymbol{i}_2; \theta)$ is the probability of correct prediction for input given the DHT parameter θ, while λ_1 and λ_2 are the hyper-parameters to balance the two loss terms.

3 Experiments

3.1 Dataset and Preprocessing

In this study, we verified the effectiveness of our proposed DHT model on the identification of infants with prenatal opioid exposure for its high risk and increasingly growing prevalence. Specifically, we used an in-house high-quality MRI dataset including 210 structural MRI scans (with both T1w and T2w images) (76 positive / 134 negative samples) acquired at different ages ranging from 6 to 439 days. The resolution of both T1w and T2w images is $0.8 \times 0.8 \times 0.8\ mm^3$. All structural MR images were processed by a state-of-the-art infant-tailored pipeline (https://www.ibeat.cloud/) [14-17], including co-registration, intensity inhomogeneity correction, skull stripping, cerebellum removal, tissue segmentation, hemispheres separation, topological correction, and surface reconstruction. Eight types of morphological features, i.e., local gyrification index, average convexity, mean curvature, sulcal depth, cortical thickness, surface area, cortical volume, and myelin content were computed as the input feature for each vertex on the cortical surface. Then, all spherical surfaces were aligned onto their age-matched templates in the 4D Infant Cortical Surface Atlas (https://www.nitrc.org/projects/infantsurfatlas/) [14, 16] and further resampled to have the same tessellation on the 6th subdivision of icosahedron with 40,962 vertices.

3.2 Experimental Settings

To validate our method, a 5-fold cross-validation strategy was employed. To quantitatively evaluate the performance, we applied four metrics to evaluate the classification performance, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under receiver operating characteristic curve (AUC), which are respectively defined as: $ACC = (TP + TN)/(TP + TN + FP + FN)$, $SEN = TP/(TP + FN)$, $SPE = TN/(TN + FP)$, where TP, TN, FP and FN are denoted as true positive, true negative, false positive and false negative values, respectively. ACC, SEN, and SPE are calculated using the default threshold of 0.5. AUC is calculated on all possible pairs of true positive rate ($TPR = SEN$) and false positive rate ($FPR = 1 - SPE$) by

changing the thresholds performed on the prediction results from our trained DHT network. In the testing phase, the mean and standard deviation of the 5-fold results were reported.

In our implementation, the feature representations produced by the five spherical transformer blocks in both $\boldsymbol{E}_{s_l}$ and $\boldsymbol{E}_{s_r}$ have 32, 32, 64, 64, and 128 channels, respectively. Correspondingly, the latent space dimension for volume transformer is 128, i.e., $C = 128$. The classifier was designed as a two-layer MLP with the ReLU activation function and the dimension of $\{128, 1\}$.

We implemented the model with PyTorch and accelerated by an NVIDIA GeForce RTX 3090 GPU. Adam was employed as optimizer with the weight decay of 10^{-4} and the learning rate was cyclically tuned within $[10^{-6}, 10^{-3}]$. The batch size was set to 16. The maximum training epoch is 500. After comparison, we empirically set the hyperparameters as λ_1=0.001 and λ_2=1.0. During the training phase, we augmented the T1w and T2w images by random erasing [18] with the probability of 0.9.

Table 1. Classification results obtained by the competing volume-based methods, surface-based methods, and our DHT. Mean and standard deviation values (mean ± std) of the testing results based on 5-fold cross validation were reported.

	Method	**ACC (%)**	**AUC (%)**	**SEN (%)**	**SPE (%)**
Volume	*ResNet3D*	73.1 ± 11.2	64.0 ± 6.0	51.7 ± 35.3	76.0 ± 8.6
	GFNet	68.9 ± 12.4	55.8 ± 8.7	51.9 ± 24.7	75.7 ± 11.0
	DAMIDL	72.5 ± 6.1	61.0 ± 5.2	51.3 ± 25.9	74.5 ± 4.0
Surface	*SUNet*	73.6 ± 3.5	67.0 ± 4.7	56.0 ± 16.4	80.1 ± 8.5
	UGSCNN	79.0 ± 4.3	67.9 ± 14.6	35.0 ± 23.1	77.5 ± 4.0
	S-Transformer	76.7 ± 2.2	71.0 ± 8.4	65.2 ± 9.4	83.4 ± 6.9
Both	***Proposed***	**80.8 ± 2.0**	**78.3 ± 9.6**	**78.6 ± 4.8**	**88.1 ± 4.5**

3.3 Results

We compared our DHT with six baseline methods, including a conventional volume-based method (i.e., ResNet3D [19]), two state-of-the-art volume-based methods for Alzheimer's disease diagnosis (i.e., GFNet [3], DIMADL [2]), and three advanced cortical surface-based methods (i.e., SUNet [11], UGSCNN [20], Spherical Transformer (S-Transformer) [4]). We implemented the competing methods based on their released code and took their encoders' output as the latent representation with a two-layer perceptron as the classifier.

As shown in **Table 1**, our DHT method achieves better performance over baselines in identifying infants with prenatal drug exposure on all evaluation metrics (i.e., ACC=80.8%, SEN=78.6%, SPE=88.1%, and AUC=78.3%). Additionally, all surface-based methods (i.e., SUNet, UGSCNN, and S-Transformer) outperform volume-based methods (i.e., ResNet3D and DAMIDL). This indicates that surface-based methods can capture more discriminative cortex features under the overwhelming dramatic brain

development during infancy. Meanwhile, among the baselines, the transformer based-method (i.e., S-Transformer) also yields better results than other five methods in most cases (i.e., SEN=65.2%, SPE=83.4%, and AUC=71.0%). The main reason could be that the transformer-based method is more suitable for the surface-based classification task by formulating the long-range dependency and leveraging the parameters more effectively and thus is less prone to overfitting with limited data size. Furthermore, compared with the advanced surface-based method UGSCNN, our DHT outperforms it by a minor margin (1.8%) in ACC. Nevertheless, our DHT has a superior improvement over UGSCNN on the sensitivity metric (SEN), which implies that our proposed method has much lower missed identification rate on the existence of prenatal drug exposure.

To further evaluate the effectiveness of the components used in our study, we further compared the proposed DHT method with its counterparts, i.e., the model with either volume data (*w/o Surface Data*) or surface data (*w/o Volume Data*), and the model without disentanglement (*w/o* $Disentanglement$). We performed *w/o* $Disentanglement$ by setting λ_1=0. As shown in **Table 2**, each component proposed in our DHT contributes to the better identification performance. For instance, our DHT with disentanglement loss $\mathcal{L}_1$ has higher accuracy than its counterpart *w/o* $Disentanglement$. These results indicate that using hybrid volume-surface data with disentanglement strategy is more effective in enhancing the discriminative features for identification of infants with prenatal drug exposure, thus achieving better performance.

Table 2. Ablation study of each component of DHT. Mean and standard deviation values (mean ± std) of the testing results based on 5-fold cross validation were reported.

Component	ACC (%)	AUC (%)	SEN (%)	SPE (%)
w/o Surface Data	74.87 ± 12.32	71.36 ± 12.95	74.34 ± 17.61	81.60 ± 10.98
w/o Volume Data	76.66 ± 2.15	71.02 ± 8.38	65.20 ± 9.40	83.41 ± 6.85
w/o $Disentanglement$	77.83 ± 4.16	73.96 ± 6.86	71.91 ± 10.16	85.96 ± 5.48
Proposed	**80.84 ± 1.96**	**78.31 ± 9.59**	**78.63 ± 4.83**	**88.14 ± 4.51**

Based on our proposed model DHT, the prenatal opioid exposure identification rate of infants is about 80%, suggesting the plausible existence of imaging biomarkers. Moreover, our method can automatically identify potential abnormal locations in both MR images and cortical surfaces for researchers and doctors to conduct further analysis. That is, our method can suggest subject-specific discriminative brain regions, including relatively informative patches on both image volumes and cortical surfaces as shown in **Fig. 2**. Specifically, in Fig. 2(a), by analyzing the attention map in the Transformer layer $\boldsymbol{O}_i$, we highlighted 50 potential discriminative volume patches with the patch size of 16 of two randomly selected subjects, which cover ~2% voxels in the whole image. It can be observed that most of selected patches are located at the deep non-cortical regions, which validates the effectiveness of our disentanglement mechanism that enforces the surface-based and volume-based representations to learn information from complementary brain regions. Moreover, we demonstrated the discriminative surface patches in Fig. 2(b) by analyzing the attention map in the Transformer layer $\boldsymbol{O}_s$, and provided the population-based importance distribution of different morphological

features through Grad-Cam [21] in Fig. 2(c), which are in line with the findings in some conventional statistical studies. For example, in Fig. 2(b), the postcentral gyrus and the lateral occipital cortex illustrated relatively high importance, while some studies have observed the developmental abnormality of these regions in children with prenatal opioid exposure [22, 23].

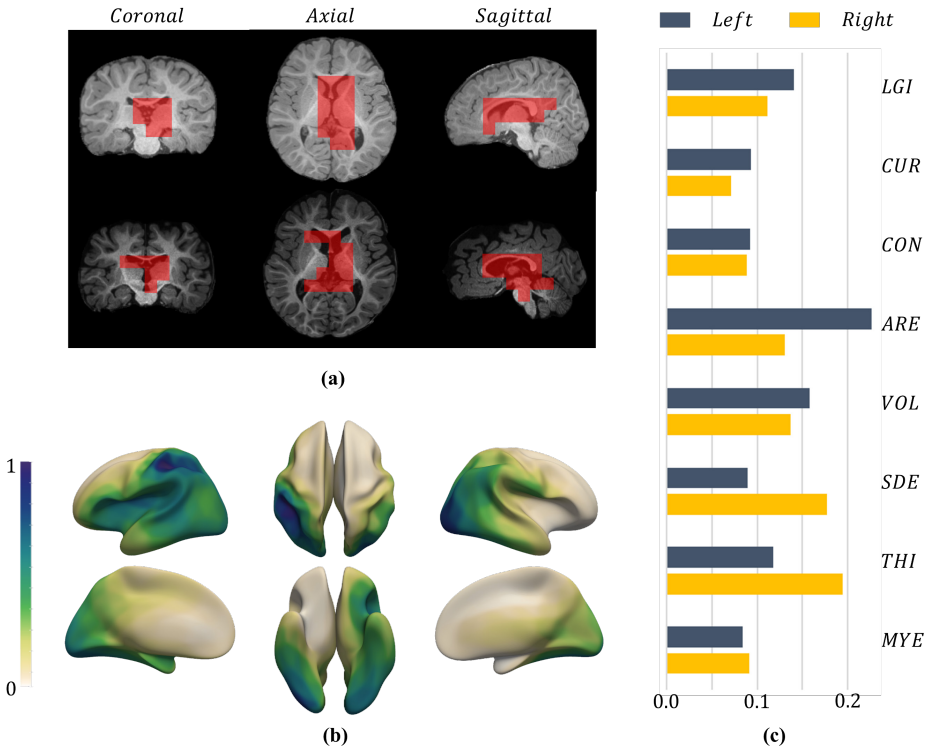

Fig. 2. The illustrations of the discriminative regions of (a) the image volumes and (b) cortical surfaces, and (c) the importance distribution of morphological features on both hemispheres. Specifically, we calculated the local gyrification index (LGI), mean curvature (CUR), surface area (ARE), volume (VOL), sulcal depth (SDE), cortical thickness (THI), and myelin content (MYE) as features for each vertex.

4 Conclusion

In this paper, we proposed the first disentangled hybrid volume-surface transformer and applied it for automatic identification of infants with prenatal drug exposure. By employing the surface-based representation for the cerebral cortex and image-based representation for deep noncortical regions, the abnormal changes affected by prenatal drug exposure can be effectively detected under the overwhelming dynamic and complicated brain development during infancy. To boost the discriminative feature

representation, we introduced the Spherical Transformer and Vision Transformer, respectively, which further embedded the hybrid data into a unified, comparable space. By disentangling their common and specific information, our DHT successfully captures the individualized representation of the infant brain for prenatal drug exposure identification. The superior identification rate compared to state-of-the-art methods demonstrates the advantages and effectiveness of our DHT, which is a powerful generic framework applicable for diagnosis of other brain disorders.

Acknowledgments. This work was supported in part by NIH grants (MH116225, MH123202, ES033518, AG075582, NS128534, and NS135574).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Morie, K. P., Crowley, M. J., Mayes, L. C., Potenza, M. N.: Prenatal drug exposure from infancy through emerging adulthood: results from neuroimaging. Drug and Alcohol Dependence **198**, 39-53 (2019).
2. Zhu, W., Sun, L., Huang, J., Han, L., Zhang, D.: Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. IEEE Transactions on Medical Imaging **40**(9), 2354-2366 (2021).
3. Zhang, S., Chen, X., Ren, B., Yang, H., Yu, Z., Zhang, X., Zhou, Y.: 3D global fourier network for Alzheimer's disease diagnosis using structural MRI. In Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 34-43 (2022).
4. Cheng, J., Zhang, X., Zhao, F., Wu, Z., Wang, Y., Huang, Y., Lin, W., Wang, L., Li, G.: Spherical transformer for quality assessment of pediatric cortical surfaces. In 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1-5 (2022).
5. Hu, D., Wang, F., Zhang, H., Wu, Z., Wang, L., Lin, W., Li, G., Shen, D., and UNC/UMN Baby Connectome Project Consortium: Disentangled intensive triplet autoencoder for infant functional connectome fingerprinting. In Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 72-82 (2020).
6. Yuan, X., Cheng, J., Zhao, F., Wu, Z., Wang, L., Lin, W., Zhang, Y., Li, G.: Multi-task joint prediction of infant cortical morphological and cognitive development. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 545-554 (2023).
7. Fischl, B., Martin I. S., and Anders M. D.: Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. NeuroImage **9**(2), 195-207 (1999).
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations (2020).
9. Cheng, J., Zhang, X., Zhao, F., Wu, Z., Yuan, X., Gilmore, J.H., Wang, L., Lin, W., Li, G.: Spherical transformer on cortical surfaces. In Proceedings of International Machine Learning in Medical Imaging Workshop, pp. 406-415 (2022).
10. Ba, J. L., Kiros, J. R., Hinton, G. E.: Layer normalization. In Advances in Neural Information Processing Systems 2016 Deep Learning Symposium (2016).

11. Zhao, F., Wu, Z., Wang, L., Lin, W., Gilmore, J.H., Xia, S., Shen, D., Li, G.: Spherical deformable u-net: Application to cortical surface parcellation and development prediction. IEEE Transactions on Medical Imaging, **40**(4), 1217-1228 (2021).
12. Zhao, F., Xia, S., Wu, Z., Duan, D., Wang, L., Lin, W., Gilmore, J.H., Shen, D., Li, G.: Spherical U-Net on cortical surfaces: methods and applications. In Information Processing in Medical Imaging: 26th International Conference (IPMI 2019), pp. 855-866 (2019).
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L, Gomez, A.N., Kaiser, L, Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems, 30 (2017).
14. Li, G., Wang, L., Shi, F., Gilmore, J.H., Lin, W., Shen, D.: Construction of 4D high-definition cortical surface atlases of infants: Methods and applications. Medical Image Analysis **25**(1), 22-36 (2015).
15. Li, G., Wang, L., Yap, P.T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, P., Shi, F., Rekik, I., Lin, W., Shen, D.: Computational neuroanatomy of baby brains: A review. NeuroImage **185**, 906-925 (2019).
16. Wu, Z., Li, W., Lin, W., Gilmore, J.H., Li, G., Shen, D.: Construction of 4D infant cortical surface atlases with sharp folding patterns via spherical patch-based group-wise sparse representation. Human Brain Mapping **40**(13), 3860-3880 (2019).
17. Wang, L., Wu, Z., Chen, L., Sun, Y., Lin, W., Li, G.: iBEAT V2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. Nature Protocols, **18**, 1488–1509 (2023).
18. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence **34**(7), pp. 13001-13008 (2020).
19. He, K., Zhang, X., Ren, S., Sun, J: Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778 (2016).
20. Jiang, C., Huang, J., Kashinath, K., Marcus, P., Niessner, M.: Spherical CNNs on unstructured grids. In International Conference on Learning Representations (2019).
21. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, pp. 618-626 (2017).
22. Radhakrishnan, R., Vishnubhotla, R.V., Guckien, Z., Zhao, Y., Sokol, G.M., Haas, D.M., Sahasivam, S.: Thalamocortical functional connectivity in infants with prenatal opioid exposure correlates with severity of neonatal opioid withdrawal syndrome. Neuroradiology, **64**(8), 1649-1659 (2022).
23. Merhar, S. L., Kline, J. E., Braimah, A., Kline-Fath, B., Tkach, J. A., Mekibib, A., He, L., Parikh, N. A.: Prenatal opioid exposure is associated with smaller brain volumes in multiple regions. Pediatric Research, **90**(2), 397-402 (2021).