2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Hybrid Learning Pipeline for Automated Diagnosis of First-Episode Schizophrenia Utilizing T1-weighted Images

Jiewei Wu[1,2], Guiwen Lyu[3], Kai Wang[1,*], Xiaoying Tang[2,*]

Abstract— **In this work, we proposed and validated a hybrid learning pipeline for automated diagnosis of first-episode schizophrenia (FES) utilizing T1-weighted images. Amygdalar and hippocampal shape abnormalities in FES have been observed in previous studies. In this work, we jointly made use of two types of features, together with advanced machine learning techniques, for an automated discrimination of FES and healthy control (96 versus 102). Specifically, we first employed a ResNet34 model to extract convolutional neural network (CNN) features. We then combined these CNN features with shape features of the bilateral hippocampi and the bilateral amygdalas, before being inputted to advanced classification algorithms such as the Gradient Boosting Decision Tree (GBDT) for classifying between FES and healthy control. Shape features were represented using log Jacobian determinants, through a well-established statistical shape analysis pipeline. When combining CNN with hippocampal shape, the best results came from utilizing GBDT as the classifier, with an overall accuracy of 75.15%, a sensitivity of 69.35%, a specificity of 80.19%, an F1 of 72.16%, and an AUC of 79.68%. When combing CNN and amygdalar shape, the best results came from utilizing Bagging as the classifier, with an overall accuracy of 74.39%, a sensitivity of 67.93%, a specificity of 80%, an F1 of 71.11%, and an AUC of 80.98%. Compared with using each single set of features, either CNN or shape, significant improvements have been observed, in terms of FES discrimination. To the best of our knowledge, this is the first work that has tried to combine CNN features and hippocampal/amygdalar shape features for automated FES identification.**

Clinical relevance— **This work provides a practical method for automated diagnosis of FES based on T1-weighted images.**

I. INTRODUCTION

Schizophrenia is a chronic and severe mental disorder affecting about 20 million people worldwide [1]. It is characterized by positive symptoms such as delusions and hallucinations and negative symptoms such as blunting of affect and passive withdrawal [2]. First-episode schizophrenia (FES) represents an early stage in the neuropathology of schizophrenia. It typically occurs in the late teenage years or the early twenties. Studying FES patients provides an opportunity to better understand the progressive changes in schizophrenia and identify potential biomarkers for schizophrenia diagnosis.

Previous structural magnetic resonance imaging (MRI) studies have reported morphometric abnormalities of the bilateral amygdalas and hippocampi in FES [3], [4]. For example, Narr et al. observed significant bilateral hippocampal volume reductions in FES compared to healthy control (HC) counterparts [3]. In one of our previous studies, we found significant global volume reductions and localized surface atrophies in the bilateral amygdalas and hippocampi [4]. These studies may indicate that the morphometric abnormalities of the bilateral hippocampi and amygdalas have the capability of distinguishing FES from HC, especially the shape features since they are more sophisticated and refined than volume features and have revealed superior discrimination ability in other types of brain diseases [5], [6].

On the other hand, convolutional neural networks (CNNs) have recently shown great potential in automated diagnoses of brain disorders, such as Alzheimer's disease, Parkinson's disease, autism spectrum disorder, and schizophrenia [7]. CNN can automatically learn feature representation through a backpropagation algorithm. ResNet [8], an efficient CNN architecture, has been widely employed as a backbone for feature extraction in many recently proposed networks. Utilizing ResNet as a feature extractor may help identify potentially useful features that are difficult to be manually-crafted.

The aim of this study is two-fold. Firstly, we aim to investigate the power of the hippocampal and amygdalar shape (quantitatively represented by log Jacobian determinants) in both hemispheres in discriminating between FES and HC. Secondly, we hypothesize that adding CNN generated features could enhance the classification performance. Extensive validation experiments are conducted.

II. MATERIALS AND METHODS

A. Subjects and MRI data acquisition

A total of 198 subjects participated in this study, including 92 FES subjects (50 females, 42 males, average age: 22.40 $\pm$ 5.59 years) and 106 HC subjects (47 females, 59 males, average age: 23.68 $\pm$ 4.04 years). Exclusion criteria included: (1) history of substance abuse or dependence; (2) significant systemic or neurologic illness as assessed by clinical evaluations and medical records; (3) comorbid affective illness or schizoaffective disorder. This study was approved by the First Affiliated Hospital of Shenzhen University Ethics

This study was supported by the National Natural Science Foundation of China (62071210), the Shenzhen Basic Research Program (JCYJ20-190809120205578), the National Key R&D Program of China (2017YFC-0112404), and the High-level University Fund (G02236002).

*Correspondence to Dr. Kai Wang (wangkai23@mail.sysu.edu.cn) and Dr. Xiaoying Tang (tangxy@sustech.edu.cn)

[1]School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China

[2]Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

[3]Department of Radiology, The First Affiliated Hospital of Shenzhen University, Shenzhen, China

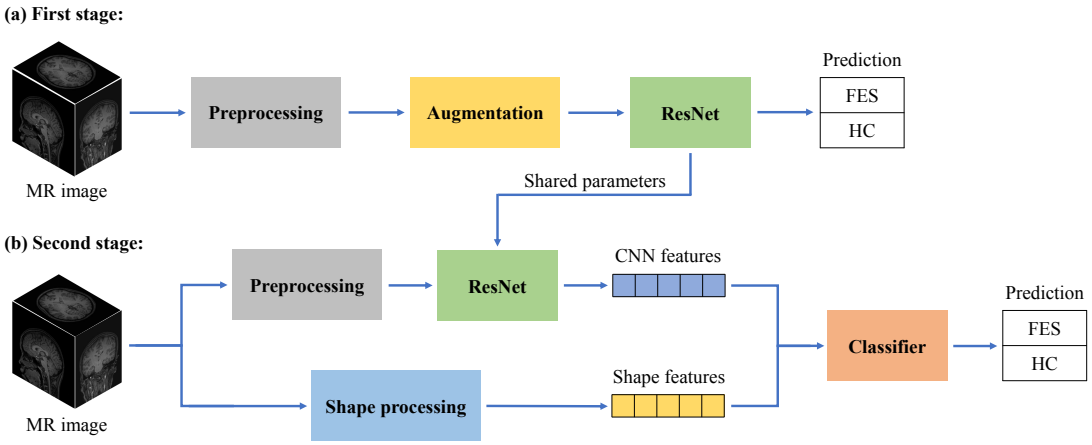

Fig. 1. The entire procedure of our proposed hybrid learning pipeline.

Committee, and was in accordance with the Declaration of Helsinki. Written informed consents were obtained from all participants or family relatives before participation.

Structural MR images were acquired using a 3T scanner (Trio Tim; Siemens, Erlangen, German). All T1-weighted images were acquired using a magnetization prepared-rapid acquisition gradient echo (MPRAGE) sequence with the following parameters: repetition time = 13.40 ms, echo time = 4.6 ms, flip angle = 20°, field of view = 256 × 256, and voxel size = 1 × 1 × 1 mm^3. Each MR image was visually inspected by one experienced neuroradiologist for data quality control.

B. Image preprocessing

All T1-weighted images in this study were processed using a specific pipeline, including bias correction, denoising, affine registration to the Colin27 MNI template [9], skull stripping. All operations were performed using the FSL package [10].

C. Volumetric segmentation

For segmenting the bilateral amygdalas and hippocampi, each T1-weighted image was processed with a validated fully automated segmentation pipeline known as *braingps* (hosted on www.mricloud.org), which is built on the multi-atlas likelihood fusion (MALF) algorithm [11]. Since MALF depends on multiple atlases, 45 atlases were used in this study. Each atlas had previously been segmented into a total of 289 regions, including our four structures of interest (i.e., left and right amygdala and hippocampus). More detailed information about the label definitions and the atlas set can be found elsewhere [12].

D. Shape processing

After segmenting out the left and right amygdala and hippocampus, a well-established shape analysis pipeline was used to extract the shape descriptor of each structure [13]. This pipeline had been successfully applied to analyzing the bilateral hippocampi and amygdalas in various brain disorders, e.g., Alzheimer's disease [5] and Wilson's disease [14]. In brief, we firstly created a triangulated surface by contouring the boundary of each 3D segmentation of each structure with sufficient smoothness and correct anatomical topology [13]. To alleviate the potential limitation of using a single surface as the template surface, we then generated a common structure-specific template surface from all 198 subjects via a Bayesian template estimation algorithm [15]. We subsequently applied the large deformation diffeomorphic metric mapping (LDDMM) surface algorithm [16] to obtain a diffeomorphism from the template to each subject surface for each structure of interest. The log Jacobian determinant of each diffeomorphism was then obtained at each vertex of the template surface for each of the four structures of interest. This diffeomorphic marker quantifies the surface deformation, namely the ratio of each subject surface to the template surface in respect of vertex-wise surface areas, and was used as our shape features in subsequent discriminant analyses.

E. Two-stage hybrid learning

As shown in Fig. 1, our proposed pipeline consists of two stages. In the first stage, the training images were randomly split into inner training images and validation images at a ratio of 3:1. A ResNet model was trained using the inner training images, and was evaluated using the inner validation images. For generating a more generalized model, online data augmentation was employed, including random cropping to a size of 172 × 206 × 172 and random scaling in the [0.95, 1.05] range. Moreover, early stopping was applied to alleviate the overfitting problem. In other words, the training process would stop when the validation loss did not decline in 25 epochs. The optimal network parameters would be saved. In the second stage, the ResNet model was initialized with the above parameters, and was then used to extract CNN features from training images and test images. This model here was only used for extracting CNN features. The shape features (the vertex-wise log Jacobian determinants) of training images and test images were extracted using the aforementioned LDDMM surface pipeline. These two types of features were extracted separately and then concatenated

TABLE I

ALL CLASSIFICATION RESULTS OF RESNET MODELS (MEAN AND STANDARD DEVIATIONS)

Model	Accuracy/%	Sensitivity/%	Specificity/%	F1/%	AUC/%
ResNet18	70.30 ± 2.38	**73.15 ± 3.51**	67.83 ± 3.76	**69.59 ± 2.40**	**75.71 ± 1.51**
ResNet34	69.14 ± 1.38	70.11 ± 3.55	68.30 ± 4.83	67.84 ± 1.11	72.02 ± 1.40
ResNet50	**71.06 ± 1.96**	70.76 ± 5.66	**71.32 ± 3.68**	69.34 ± 2.98	74.10 ± 2.01

Values expressed as mean ± standard deviation. Bold indicates the best performance. AUC, area under the receiver operating characteristic curve.

TABLE II

A SUMMARY OF ALL CLASSIFICATION RESULTS (MEAN AND STANDARD DEVIATIONS)

Shape feature	CNN feature	Classifier	Accuracy/%	Sensitivity/%	Specificity/%	F1/%	AUC/%
Amygdala	None	Bagging	71.31 ± 2.12	67.07 ± 3.19	75.00 ± 2.16	68.46 ± 2.53	77.64 ± 1.37
		GBDT	73.23 ± 2.38	67.50 ± 2.97	78.21 ± 2.81	70.08 ± 2.66	78.60 ± 2.05
Hippocampus	None	Bagging	67.12 ± 2.10	61.09 ± 3.29	72.36 ± 2.46	63.30 ± 2.57	71.83 ± 1.69
		GBDT	66.72 ± 1.54	60.43 ± 3.08	72.17 ± 1.85	62.76 ± 2.14	70.98 ± 2.19
None	ResNet18	Bagging	71.87 ± 2.37	64.13 ± 3.07	78.58 ± 3.43	67.93 ± 2.65	75.75 ± 2.48
		GBDT	70.45 ± 2.16	64.24 ± 2.85	75.85 ± 3.13	66.89 ± 2.37	74.03 ± 2.38
	ResNet34	Bagging	72.88 ± 2.03	63.91 ± 3.43	80.66 ± 2.93	68.63 ± 2.49	75.56 ± 2.55
		GBDT	70.25 ± 2.58	62.83 ± 3.94	76.70 ± 3.35	66.22 ± 3.13	74.77 ± 3.40
	ResNet50	Bagging	72.78 ± 1.70	65.11 ± 2.85	79.43 ± 2.06	68.95 ± 2.14	75.23 ± 1.75
		GBDT	71.41 ± 1.75	65.76 ± 2.89	76.32 ± 2.55	68.11 ± 2.05	74.68 ± 2.47
Amygdala	ResNet18	Bagging	71.87 ± 2.57	64.24 ± 3.59	78.49 ± 3.65	67.96 ± 2.92	79.91 ± 2.06
		GBDT	71.87 ± 1.94	65.00 ± 4.06	77.83 ± 2.84	68.18 ± 2.58	77.60 ± 2.37
	ResNet34	Bagging	74.39 ± 1.72	67.93 ± 3.31	80.00 ± 1.57	71.11 ± 2.27	80.98 ± 1.89
		GBDT	72.58 ± 2.24	66.52 ± 4.83	77.83 ± 1.42	69.19 ± 3.16	79.47 ± 2.48
	ResNet50	Bagging	72.07 ± 1.71	67.07 ± 2.75	76.42 ± 2.80	69.04 ± 1.96	79.65 ± 1.31
		GBDT	70.96 ± 1.16	66.85 ± 3.12	74.53 ± 2.92	68.12 ± 1.52	77.83 ± 1.40
Hippocampus	ResNet18	Bagging	73.64 ± 1.87	66.96 ± 3.74	79.43 ± 3.07	70.20 ± 2.38	77.99 ± 2.37
		GBDT	73.59 ± 2.56	68.59 ± 3.45	77.92 ± 4.29	70.70 ± 2.65	77.63 ± 2.79
	ResNet34	Bagging	74.65 ± 1.66	67.72 ± 3.04	**80.66 ± 2.57**	71.26 ± 2.01	79.02 ± 1.23
		GBDT	**75.15 ± 0.90**	69.35 ± 2.22	80.19 ± 2.27	**72.16 ± 1.06**	79.68 ± 1.25
	ResNet50	Bagging	72.68 ± 1.96	66.74 ± 3.04	77.83 ± 2.57	69.40 ± 2.34	78.55 ± 1.85
		GBDT	72.32 ± 2.42	67.61 ± 4.06	76.42 ± 4.74	69.40 ± 2.55	78.12 ± 2.23
Amygdala + Hippocampus	ResNet18	Bagging	73.23 ± 2.42	66.63 ± 4.27	78.96 ± 2.98	69.77 ± 3.08	80.64 ± 2.28
		GBDT	73.54 ± 2.32	68.15 ± 3.67	78.21 ± 3.71	70.51 ± 2.63	80.39 ± 2.18
	ResNet34	Bagging	74.60 ± 1.86	67.83 ± 3.00	80.47 ± 2.83	71.26 ± 2.16	81.77 ± 1.74
		GBDT	74.75 ± 1.60	**70.11 ± 2.97**	78.77 ± 3.46	72.06 ± 1.69	**81.91 ± 1.90**
	ResNet50	Bagging	73.18 ± 1.45	66.96 ± 3.08	78.58 ± 1.89	69.85 ± 2.03	80.40 ± 1.77
		GBDT	73.23 ± 1.66	69.46 ± 4.05	76.51 ± 2.88	70.64 ± 2.30	80.65 ± 2.07

Values expressed as mean ± standard deviation. Bold indicates the best performance. AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; GBDT, Gradient Boosting Decision Tree.

before classification. Please note that both training features and test features were normalized using z-score standardization. We then performed inner cross-validation on the training data to exhaustly search the optimal set of hyper-parameters from pre-defined parameter candidates using grid search strategy. The parameters of the highest accuracy were chosen to be the optimal set of hyper-parameters, which would be applied in the training process later. After the training process, we evaluated the trained model using the test data.

F. Cross-validation

In our experiments, we repeated stratified 5-fold cross-validation for 10 times. The data was randomly split into 5 folds at each time so that the partition results vary from time to time. Within each iteration, four folds were used for training and one fold for test. The proposed pipeline was repeated 5 times, and cross-validation results of the 5 folds were calculated. Lastly, the mean and standard deviation, averaged across the 10 times, were obtained.

III. EXPERIMENTS AND RESULTS

A. Experimental settings

To demonstrate the superiority of our proposed pipeline, we compared the classification performance of utilizing shape features, CNN features, and a combination of the two types of features. In our experiments, we also compared the classification power of features generated by three variants

of ResNet, namely ResNet18, ResNet34, and ResNet50. Moreover, two classifiers were evaluated, namely Bagging [17] and Gradient Boosting Decision Tree (GBDT) [18]. Accuracy, sensitivity, specificity, F1, and area under the receiver operating characteristic curve (AUC) were used as metrics to evaluate the classification performance. All experiments in this study were performed using scikit-learn [19] and PyTorch packages.

B. Cross-validation results

Table I shows the results of all ResNet experiments. Clearly, ResNet50 achieved the highest overall accuracy of 71.06%. This may be because it has a relatively stronger representation ability than other models. However, generally speaking, ResNet models alone did not provide satisfactory results, which might be due to the limited data size.

The results of all hybrid learning experiments are shown in Table II. When combined with the amygdalar shape features, the hippocampal shape features, or a combination of the two, ResNet34 consistently worked the best compared to the other two ResNet variants. With GBDT being the classifier, an overall accuracy of 75.15%, a sensitivity of 69.35%, a specificity of 80.19%, an F1 of 72.16%, and an AUC of 79.68% were achieved when using a combination of hippocampal shape features and ResNet34 features. With Bagging being the classifier, an overall accuracy of 74.39%, a sensitivity of 67.93%, a specificity of 80%, an F1 of 71.11%, and an AUC of 80.98% were obtained when using a combination of amygdalar shape features and ResNet34 features. With GBDT being the classifier, an overall accuracy of 74.75%, a sensitivity of 70.11%, a specificity of 78.77%, an F1 of 72.06%, and an AUC of 81.91% were obtained when using a combination of amygdalar shape features, hippocampal shape features and ResNet34 features. Compared with using each single set of features, significant improvements were observed, especially in terms of accuracy and AUC. This clearly indicates that CNN features and shape features of the bilateral hippocampi as well as the bilateral amygdalas provide complementary information for each other, in terms of FES discrimination.

There are several other interesting findings that are worthy of mentioning. Firstly, compared with using the hippocampal shape features only, adding ResNet50 features did not improve the classification performance as much as ResNet34 features did. We conjecture a plausible reason is that the dimension of ResNet50 features (2048) was 4 times of that of ResNet34 features (512), and is likely to induce overfitting. Secondly, in some experimental settings, adding ResNet features did not improve the classification performance compared with using the amygdalar shape features only, which probably because these two types of features shared redundant information.

IV. CONCLUSIONS

In this study, we proposed and validated a hybrid learning pipeline for automated classification of FES. Our experiments demonstrate that a combination of hippocampal or amygdalar shape features and ResNet34-generated features could improve the FES identification performance. A potential limitation of this work is that the ResNet model was frozen after the first stage and could not be updated during the second stage. In the future, we aim to design a deep hybrid learning pipeline, which can backpropagate and update the parameters of the CNN module. Another limitation is that we have not applied our models in clinical practice, which will be one of our future endeavors.

REFERENCES

[1] G. Collaborators et al., Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. 2018.

[2] S. R. Kay, A. Fiszbein, and L. A. Opler, The positive and negative syndrome scale (panss) for schizophrenia, Schizophrenia bulletin, vol. 13, no. 2, pp. 261–276, 1987.

[3] K. L. Narr, P. M. Thompson, P. Szeszko, D. Robinson et al., Regional specificity of hippocampal volume reductions in first-episode schizophrenia, Neuroimage, vol. 21, no. 4, pp. 1563–1575, 2004.

[4] X. Tang, G. Lyu, M. Chen, W. Huang et al., Amygdalar and hippocampal morphometry abnormalities in first-episode schizophrenia using deformation-based shape analysis, Frontiers in Psychiatry, vol. 11, p. 677, 2020.

[5] X. Tang, D. Holland, A. M. Dale, L. Younes et al., Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer's disease: detecting, quantifying, and predicting, Human brain mapping, vol. 35, no. 8, pp. 3701–3725, 2014.

[6] X. Tang, D. Holland, A. M. Dale, L. Younes et al., Baseline shape diffeomorphometry patterns of subcortical and ventricular structures in predicting conversion of mild cognitive impairment to alzheimer's disease, Journal of Alzheimer's Disease, vol. 44, no. 2, pp. 599–611, 2015.

[7] L. Zhang, M. Wang, M. Liu, and D. Zhang, A survey on deep learning for neuroimaging-based brain disorder analysis, Frontiers in neuroscience, vol. 14, 2020.

[8] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[9] C. J. Holmes, R. Hoge, L. Collins, R. Woods et al., Enhancement of mr images using registration for signal averaging, Journal of computer assisted tomography, vol. 22, no. 2, pp. 324–333, 1998.

[10] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich et al., Fsl, Neuroimage, vol. 62, no. 2, pp. 782–790, 2012.

[11] X. Tang, K. Oishi, A. V. Faria, A. E. Hillis et al., Bayesian parameter estimation and segmentation in the multi-atlas random orbit model, PloS one, vol. 8, no. 6, p. e65591, 2013.

[12] D. Wu, C. Ceritoglu, M. I. Miller, and S. Mori, Direct estimation of patient attributes from anatomical mri based on multi-atlas voting, NeuroImage: Clinical, vol. 12, pp. 570–581, 2016.

[13] X. Tang, Y. Luo, Z. Chen, N. Huang et al., A fully-automated subcortical and ventricular shape generation pipeline preserving smoothness and anatomical topology, Frontiers in neuroscience, vol. 12, p. 321, 2018.

[14] L. Zou, Y. Song, X. Zhou, J. Chu et al., Regional morphometric abnormalities and clinical relevance in wilson's disease, Movement Disorders, vol. 34, no. 4, pp. 545–554, 2019.

[15] J. Ma, M. I. Miller, and L. Younes, A bayesian generative model for surface template estimation, International journal of biomedical imaging, vol. 2010, 2010.

[16] M. Vaillant and J. Glaunes, Surface matching via currents, in Biennial International Conference on Information Processing in Medical Imaging, Springer, pp. 381–392, 2005.

[17] L. Breiman, Bagging predictors, Machine learning, vol. 24, no. 2, pp. 123–140, 1996.

[18] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics, pp. 1189–1232, 2001.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel et al., Scikit-learn: Machine learning in python, the Journal of machine Learning research, vol. 12, pp. 2825–2830, 2011.