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

Self-Paced Learning and Privileged Information based Cascaded Multi-column Classification algorithm for ASD diagnosis

Yu Zhang, Bo Peng, Zeyu Xue, Jian Bao, Bing Keong Li, Yan Liu, Yuqi Liu, Mao Sheng, Chunying Pang*, Yakang Dai*

Abstract—**Autism spectrum disorder (ASD) is one of the most serious mental disorder in children. Machine learning based computer aided diagnosis (CAD) on resting-state functional magnetic resonance imaging (rs-fMRI) for ASD has attracted widespread attention. In recent years, learning using privileged information (LUPI), a supervised transfer learning method, has been generally used on multi-modality cases, which can transfer knowledge from source domain to target domain in order to improve the prediction capability on the target domain. However, multi-modality data is difficult to collect in clinical cases. LUPI method without introducing additional imaging modality images is worth further study. Random vector function link network plus (RVFL+) is a LUPI diagnosis algorithm, which has been proven to be effective for classification tasks. In this work, we proposed a self-paced learning based cascaded multi-column RVFL+ algorithm (SPL-cmcRVFL+) for ASD diagnosis. Initial classification model is trained using RVFL on the single-modal data (e.g. rs-fMRI). The output of the initial layer is then sent as privileged information (PI) to train the next layer of classification model. During this process, samples are selected using self-paced learning (SPL), which can adaptively select simple to difficult samples according to the loss value. The procedure is repeated until all samples are included. Experimental results show that our proposed method can accurately identify ASD and normal control, and outperforms other methods by a relatively higher classification accuracy.**

I. Introduction

Autism spectrum disorder (ASD) is a kind of complex neurodevelopmental disorders that can affect patients' social behavioral and communication abilities. Traditionally, diagnosis of ASD depends mainly on neuropsychiatric experts, which can easily be misdiagnosed.

Presently, most studies have focused on using magnetic resonance imaging (MRI) to detect abnormalities for ASD [1].Specifically, functional MRI (fMRI) is a useful tool [2]. Plitt *et al.* shown that highly predictive brain features to diagnose ASD can be obtained based on resting-state fMRI (rs-fMRI) [5]. In addition, the computer-aided diagnosis (CAD) for ASD based on MRI has also received widespread attention. In the whole CAD system, the classifier is the key component, which involves the classification performance.

In recent years, a variety of classifiers based on learning using privileged information (LUPI) have been proposed and used in the field of transfer learning. For example, Vapnik *et al*. proposed a privileged information (PI) learning paradigm Support Vector Machine plus (SVM+) [6], noting that PI only exists in the training process. Zhang *et al.* proposed a random vector functional link network plus (RVFL+) [7], which has better performance than SVM+. In addition, Shi *et al.* proposed a cascaded multi column RVFL+ (cmcRVFL+) framework [8], which achieved good classification performance for Parkinson's disease diagnosis. However, this method did not consider the impact of sample selection in training of the classifier.

Self-paced learning (SPL) is an extension of course learning that is more flexible and adaptive in estimating the difficulty of sample learning [9]. The core idea of SPL is to select samples with small training errors from all samples during each iteration, which is used to update the model parameters according to certain criteria. Zhang *et al.* proposed a weakly supervised training framework based on self-paced learning for vessel segmentation in X-ray angiography [12]. Zhu *et al.* proposed a multi-modal rank minimization method based on self-paced learning for Alzheimer's disease classification [13]. However, the above research work does not involve privileged information learning and cascaded classifier.

In this paper, we propose a self-paced learning cascaded multi-column RVFL+ algorithm (SPL-cmcRVFL+) based on rs-fMRI for ASD diagnosis. Specifically, in the first iteration, the initial proportion of samples are sent to the cascaded multi-column training network, and the predictive label values of the front layer is used as the PI of the next layer for auxiliary training, where each layer integrates multi-column classifiers. SPL evaluates the loss of samples value from the last layer classifier, introducing new samples according to proportional attenuation, which is subsequently used for the next iteration of training. The iteration process is completed only when all samples are included in the training.

* This work is partly supported by the National Natural Science Foundation of China (61971413, 61801476), Jiangsu Natural Science Foundation (BE2021780051, BK20180221), Shandong Natural Science Foundation (ZR2020QF022), Jinan Innovation Team (2018GXRC017), Jilin Science & Technology Department Project (20200404212YY), and Suzhou Science & Technology Projects for People's Livelihood (SS202065)

Y. Zhang and C. Pang is with School of Life Science and Technology, Changchun University of Science and Technology, Changchun, China (e-mail: pangcy@cust.edu.cn).

B. Peng, Z. Xue, Y. Liu and Y. Dai is with Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China, and with Jinan Guoke Medical Engineering Technology Development co., LTD, Jinan, China (e-mail: daiyk@sibet.ac.cn).

B. K. Li and J. Bao is with Jiangsu LiCi Medical Device Co., Ltd., Lianyungang, China.

Y. Liu and M. Sheng is with Department of Radiology, Children's Hospital of Soochow University, Suzhou, China.

II. METHOD

A. The algorithm framework

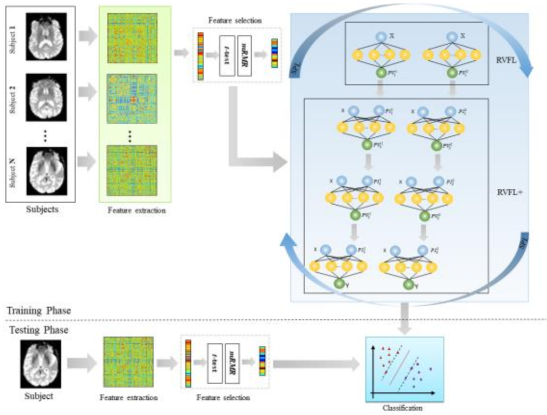

Figure 1. Flowchart of the proposed SPL-cmcRVFL+ algorithm

The flow chart of the proposed SPL-cmcRVFL+ is shown in Figure 1. The steps involved for the algorithm are as follows:

(1) The features from the brain rs-fMRI images are first extracted.

(2) The t-test method and the minimum redundancy maximum correlation method (mRMR) are used to select the features.

(3) RVFL classifier with multiple columns is used to initialize the classifier as the first layer. The predictive label values are used as PI to input into the second-layer RVFL+ classifier with multiple columns. We repeat this process to build a cascaded multi-column training network. It is worth noting that the RVFL+ parameters of each layer are fine-tuned by the new PI value.

(4) In the training of the classifier, SPL is used to adaptively select samples according to the learning difficulty from simple to difficult according to the loss value, until all samples are included or the loss function can no longer be reduced.

(5) The final classifier is obtained after the final RVFL+ results are obtained. 5-fold cross validation is conducted to test the effect of the model in the training phase.

(6) In the testing stage, the testing samples without sample selection are directly sent to the trained network.

B. RVFL+

The objective function of RVFL+ is defined as:

$$\min_{\beta,\tilde{\beta}} \frac{1}{2}\|\beta\|_2^2 + \frac{\varepsilon}{2}\left\|\tilde{\beta}\right\|_2^2 + C\sum_{i=1}^{n} \check{h}(p_i)\tilde{\beta}$$

$$\text{s.t. } h(x_i)\beta = y_i - \mu_i\left(\tilde{\beta}, \check{h}(p_i)\right), \forall 1 \le i \le n \quad (1)$$

where p_i is the PI, ε is a regularization coefficient, μ_i is a training error vector, $\check{h}(p_i)$ is the connection vector of the privileged information space in the same form as $h(x_i)$, $h(x_i)$ represents feature matrix, and the $\mu_i\left(\tilde{\beta}, \check{h}(p_i)\right)$ is the correcting function with the output weight vector $\tilde{\beta}$ in the privileged feature space. The details of RVFL+ algorithm please refer to [8].

In order to solve the optimization problem of Eq. (1), the Lagrangian function is defined as

$$\min_{\beta,\tilde{\beta}} \frac{1}{2}\|\beta\|_2^2 + \frac{\varepsilon}{2}\left\|\tilde{\beta}\right\|_2^2 + C\sum_{i=1}^{n} \check{h}(p_i)\tilde{\beta}$$

$$-\sum_{i=1}^{n} \lambda_i\left(h(x_i)\beta - y_i + \check{h}(p_i)\tilde{\beta}\right) \quad (2)$$

where $\lambda = [\lambda_1, \cdots \lambda_n]^T$ are Lagrange multipliers. Furthermore, by using the Karush-Kuhn-Tucker condition [14] to calculate the saddle points, the output function of the RVFL+ is finally given as

$$f(x) = h(x)\beta =$$

$$h(x)H^T\left(HH^T + \frac{1}{\varepsilon}\tilde{H}\tilde{H}^T + \frac{1}{C}\right)^{-1}\left(Y - \frac{C1}{\varepsilon}\tilde{H}\tilde{H}^T\right) \quad (3)$$

C. SPL-cmcRVFL+

SPL introduces a binary variable v_i to control whether eac h sample is selected in the traditional machine learning object ive function. The objective function is as follow

$$w_{t+1} = argmin(r(w) + \sum_{i=1}^{n} f(x_i, y_i, w)) \quad (4)$$

where the regularization term is $r(w)$, and the loss function of the model is $f(x_i, y_i, w)$. Only when v_i=1 can the sample be included in the calculation of the objective function, and the objective function can be changed to

$$(w_{t+1}, v_{t+1}) =$$

$$argmin\left(r(w) + \sum_{i=1}^{n} v_i f(x_i, y_i, w) - \frac{1}{K}\sum_{i=1}^{n} v_i\right) \quad (5)$$

where K is used to control the number of samples. If K value is large, the number of non-zero v_i is allowed to be small.The sample with small loss value are selected in the objective function optimization process. The optimization process is defined as

$$v_i = \delta\left(f\left(x_i, y_i, w\right) < \frac{1}{K}\right) \quad (6)$$

Under the framework of our proposed SPL-cmcRVFL+ algorithm, the first iterative training sample set $D_1 = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$ is obtained through the K value mentioned above. These training samples, which are less than the number of original samples set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l), \cdots, (x_n, y_n)\}$, are fed to the first layer of multi-column RVFL classifier for training, and the generated PI is sent into the next layer of RVFL+ classifier. A new training sample set D_2 is generated and used in the next iteration of training. This iteration ends until the last layer of classifier training is completed when all samples are selected or the loss function can no longer be reduced.

III. EXPERIMENT AND RESULTS

A. Dataset and Imaging processing

In this study, 50 subjects (23 ASD patients and 27 normal controls) are obtained from Autism Brain Imaging Data Exchange (ABIDE) initiative dataset (http://fcon_1000.projects.nitrc.org/indi/abide) .The classification performance is evaluate on the dataset.

TABLE 1. PERFOEMANCE COMPARISON WITH DIFFERERNT ALGORITHM(UNIT:%)

	ACC	SPC	SEN	AUC
ELM	81.67±2.46	83.01±8.45	83.97±6.30	53.61
RVFL	83.72±6.24	85.11±13.91	82.66±19.20	69.14
cmcRVFL+ (7cols)	82.64±10.21	82.66±2.78	82.33±22.4	68.02
cmcRVFL+ (5cols)	81.16±10.55	78.66±6.91	84±26.07	66.91
cmcRVFL+ (3cols)	82.51±9.65	82.44±3.88	82.33±25.04	68.56
cmcRVFL+ (2cols)	83.85±11.62	83.55±7.83	84±26.07	69.51
SPL-cmcRVFL+ (2cols)	89.71±7.53	92±10.95	87.55±11.39	76.40

The rs-fMRI data are preprocessed using DPABI (http://rfmri.org/dpabi) toolkit [15]. The steps include: (1) Format conversion, (2) Discard the first 10 volumes, (3) Slice timing to correct temporal difference in acquisition among different slices, (4) Realign to correct for head motion, (5) Normalize to Montreal Neurological Institute (MNI) space, (6) Image smoothing to reduce spatial noise, (7) Noise filtering to remove the effects of low-frequency drift and high-frequency noise. Lastly, the function connections of pairs of ROIs are computed. Due to the similarity of function connection matrix, only the elements in upper triangular matrix are included for concatenation of the elements into a vector. Through the above process, each participant correspond to a feature vector.

B. *Experiment*

To verify the effectiveness of the proposed algorithm, we compare SPL-cmcRVFL+ with three machine learning algorithms: (1) ELM; (2) RVFL; (3) cmcRVFL+ (a machine learning method similar to RVFL). In order to test the effect of the model, we adopted a 5-fold cross-validation strategy and calculated the average classification results after repeated operation. Classification accuracy (ACC), sensitivity (SEN), specific (SPC) and area under the curve (AUC) are chosen as evaluation indexes.

C. *Results*

Table 1 shows the comparison performance of other machine learning algorithms. It can be found that SPL-cmcRVFL+ outperforms other algorithms. The ACC, SEN, SPC, and AUC of SPL-cmcRVFL+ is 89.71%, 92%, 87.55% and 76.4% respectively. The proposed SPL-cmcRVFL+ has the best overall result in comparison. The classification results of SPL-cmcRVFL+ are better than those of ELM and RVFL, which shows that the combination of cascaded multi column classifiers can improve the classification performance. In comparison with cmcRVFL+, the proposed SPL-cmcRVFL+ also achieves better classification results, which shows that classification performance can be improved according to the learning difficulty, using SPL to gradually incorporate training samples into the training,.

Figure 2 shows the classification performance of four layers of cmcRVFL+ with different number of columns. It is found that the network achieves the best classification effect by integrating two columns of classifiers in each layer, which indicates the ensemble learning of classifiers can improve the classification performance.

Figure 3 shows the classification results in two columns of cmcRVFL+ with different layers. It is found that the classifier model achieves the best classification results with four layers cascaded and two columns integrated in each layer, which indicated that the output weight is continuously optimized, and the classification performance is improved with the deepening of classifier layers. Therefore, we take four-layer and two-column as classification module, and apply SPL to cmcRVFL+ to train the network in the proposed SPL-cmcRVFL+.

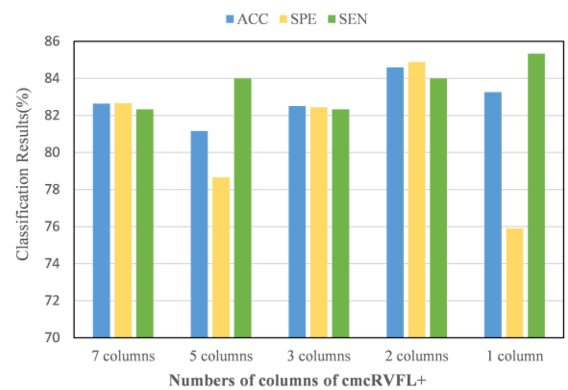

Figure 2. Classification performance of four-layer cmcRVFL+ algorithm with different columns

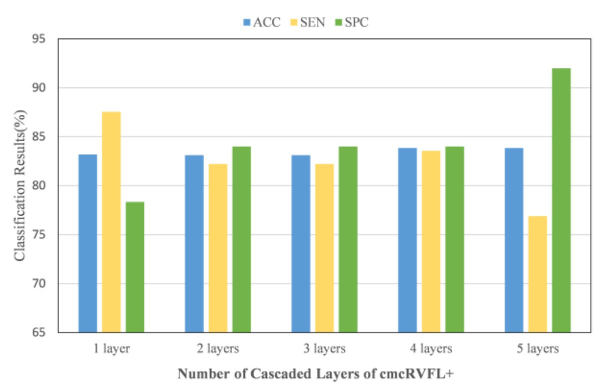

Figure 3. Classification performance of two-column cmcRVFL+ algorithm with different layers

Figure 4 shows the ROC curves of our algorithm and other compared methods. It is shown that the AUC of SPL-cmcRVFL+ algorithm is 76.4%, which indicates the proposed classifier has a high prediction accuracy.

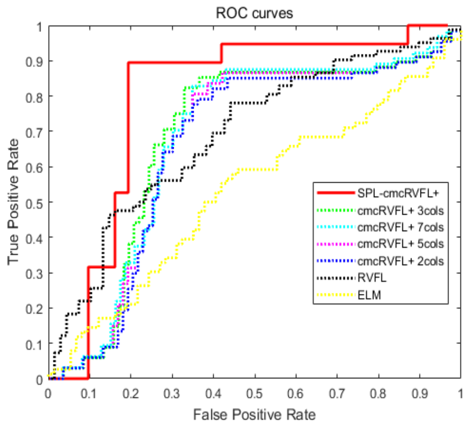

Figure 4. ROC curves for different algorithms

IV. CONCLUSION

The traditional LUPI classifier requires multimodal PI in the training stage, which limits its application. In clinical practice, not every patient receives multimodal medical imaging examination. Therefore, single-modality image based CAD is more common. A SPL-cmcRVFL+ method is proposed, the output of front layer classifier is regarded as PI for next layer, which is iterated without additional modal information. Traditional classifier modeling is training using all samples simultaneously, ignoring the influence of sample attributes on the classifier. Instead, we first used SPL algorithm to select samples with relatively small-loss value, and therefore adjust the model in iteration of training stage. This will ensure that samples with large loss (possibly noise points or outliers) will not enter into each iteration, which improves the robustness of the training model. In order to prevent the occurrence of small training error and large generalization error, regularization term is used to control the complexity of the classification model in training, which may reduce the risk of overfitting.

Although, the proposed SPL-cmcRVFL+ achieves the best classification performance compared to the competing approaches, the proposed algorithm still need to be further optimized especially in AUC for clinical application. The sample size of the public data set used in this paper is relatively small, which affects the effectiveness of the model to a certain extent. More ASD data as well as statistical analysis are required for in-depth and comprehensive testing of the algorithm on large data set.

To sum up, a cascaded multi-column LUPI classifier framework has been proposed based on SPL for assisted diagnosis of ASD with single modality neuroimaging (rs-fMRI). Compared with ELM, RVFL and cmcRVFL+ algorithm, the proposed SPL-cmcRVFL+ algorithm has better classification with an increase accuracy of 5.8~8.5%. The current proposed SPL-cmcRVFL+ algorithm is tested on diagnosis of ASD, which can also be applied to diagnose other brain disease based on single-mode image, and will be further investigated.

REFERENCES

[1] C. Ecker, S. Y. Bookheimer, and D. G. M. Murphy, "Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan," Lancet Neurology, vol. 14, pp. 1121-1134, 2015.

[2] R. Müller, P. Shih, B. Keehn, J. R. Deyoe, K. M. Leyden and D. K. Shukla, "Underconnected, but How? A Survey of Functional Connectivity MRI Studies in Autism Spectrum Disorders, "Cerebral Cortex, vol. 21, no. 10, pp. 2233–2243, 2011.

[3] M. E. Vissers, M. X. Cohen and H. M. Geurts, "Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links," Neuroscience & Biobehavioral Reviews, vol. 36, pp. 604-625, 2012.

[4] L. Q. Uddin, K. Supekar, and V. Menon, "Reconceptualizing functional brain connectivity in autism from a developmental perspective." Frontiers in Human Neuroscience, vol. 7, pp. 458, 2013.

[5] P. Mark, K. A. Barnes, and M. Alex, "Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards," Neuroimage:Clinical, vol. 7, pp. 359–366, 2015.

[6] V. Vapnik and A. Vashist, "A new learning paradigm: learning using privileged information," Neural Networks, vol. 22, pp. 544-557, 2009.

[7] P. B. Zhang and Z. Yang, "A new learning paradigm for random vector functional link network: RVFL+," Neural Networks, vol. 122, pp. 94-105, 2020

[8] J. Shi, Z. Xue, Y. Dai, B. Peng, Y. Dong, Q. Zhang and Y. Zhang, "Cascaded Multi-Column RVFL+ Classifier for Single-Modal Neuroimaging-Based Diagnosis of Parkinson's Disease," IEEE Transactions on Biomedical Engineering, vol. 66, pp. 2362-2371, 2019.

[9] L. Lin, K. Wang, D. Meng, W. Zuo and L. Zhang, "Active Self-Paced Learning for Cost-Effective and Progressive Face Identification," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, pp. 7-19, 2018.

[10] Q. Zhao, D. Meng, L. Jiang, Q. Xie, Z. Xu and A.G. Hauptmann, "Self-paced learning for matrix factorization," AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 3196–3202, 2015.

[11] D. Meng, Q. Zhao and L. Jiang, "A theoretical understanding of self-paced learning," Information Sciences, vol. 414, pp. 319-328, 2017.

[12] J. Zhang, G. Zhang, H. Xie, S. Zhang and L. Gu, "Weakly Supervised Vessel Segmentation in X-ray Angiograms by Self-Paced Learning from Noisy Labels with Suggestive Annotation," Neurocomputing, vol. 417, pp. 114-127, 2020.

[13] Q. Zhu, N. Yuan, J. Huang, X. Hao and D. Zhang, "Multi-Modal AD Classification via Self-paced Latent Correlation Analysis," Neurocomputing, vol. 355, pp. 143-154, 2019.

[14] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya and K. R. K. Murthy, "Improvements to Platt's SMO Algorithm for SVM Classifier Design," Neural Computation, vol. 13, no. 3, pp. 637–649. 2001.

[15] C. G. Yan, X. D. Wang, and X.N.Zuo, "DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging," Neuroinformatics, vol. 14, pp. 339-351, 2016.