Connecting Patients with Pre-diagnosis: A Multiple Graph Regularized Method for Mental Disorder Diagnosis

Tianqi Zhao[1], Ming Kong[1], Kun Kuang[1,4(✉)], Zhengxing Huang[1], Qiang Zhu[1], and Fei Wu[1,2,3]

[1] Institute of Artificial Intelligence, Zhejiang University, Hangzhou, China
{ztqwdk,zjukongming,kunkuang,zhengxinghuang,zhuq}@zju.edu.cn, wufei@cs.zju.edu.cn
[2] Shanghai Institute for Advanced Study of Zhejiang University, Shanghai, China
[3] Shanghai AI Laboratory, Shanghai, China
[4] Key Laboratory for Corneal Diseases Research of Zhejiang Province, Hangzhou, China

Abstract. Computer-aided diagnosis (CAD) plays an important role in medicine. But most of the previous methods only focus on the diagnosis process information like the image data for medical patterns learning, ignoring the pre-diagnosis, which is necessary and important for a doctor's decision. Besides, traditional CAD methods treat the patients as independent samples in data. To make up this gap, in this paper, we propose to connect patients with pre-diagnosis and propose a novel Multiple Graph REgularized Diagnosis (MuGRED) method for mental disorder diagnosis, which contains two main components: multi-modal representation learning and a multiple graph feature fusion module. We validated our MuGRED method on a practical dataset of children's attention deficit and hyperactivity disorder and a well-recognized ASD benchmark. Extensive experiments demonstrate that our MuGRED method can achieve a better performance than the state-of-the-art methods for mental disorder diagnosis.

Keywords: Computer-aided diagnosis · Mental disorder · Graph attention network

1 Introduction

To simplify the diagnosis process and alleviate the lack of medical resources, computer auxiliary analyses of the information generated in the process of medical diagnosis have become the focus of both academia and industry [4,12]. Most of the related works of computer-aided diagnosis focus on the doctor's diagnosis process of the patient's condition, such as analyzing the patient's medical images [5,19], or the doctor's performance when communicating with the patient [1,22]. However, most of the current computer-aided methods only focus on analyzing

L. Fang et al. (Eds.): CICAI 2022, LNAI 13605, pp. 362–374, 2022.
https://doi.org/10.1007/978-3-031-20500-2_30

the doctors' diagnosis process information for medical patterns learning, ignoring the fact that the medical diagnosis process is a comprehensive, complicated and multi-step judgment process.

Pre-diagnosis is one of the most important procedures during the diagnosis process. But how to combine the pre-diagnosis information with the computer-aided diagnosis model to obtain more accurate conclusions has not yet been paid enough attention. So we propose the *Multiple Graph Regularized Diagnosis* (MuGRED) method to generate the pre-diagnosis information with the image process diagnosis. The proposed MuGRED method consists of two main components: multimodal representation learning, which is designed to learn the representation of patients from the diagnosis process, and a multiple graph feature fusion module to fuse the information from varieties of pre-diagnosis as well as from the diagnosis process for the final diagnosis conclusion.

In this paper, we focus on the diagnosis of mental disorders. We evaluate our method's effectiveness on two datasets: a practical dataset for ADHD and a well-recognized mental disorders dataset for anxiety and depression. We introduce the pre-diagnosis results to these two multimodal behavior analysis problems and compare them with state-of-the-art methods. Experimental results show that our method can better use pre-diagnosis information to achieve a more accurate diagnosis. The main contributions of this work can be summarized as follows:

- We propose a novel end-to-end *Multiple Graph Regularized Diagnosis* model for mental disorder diagnosis. To the best of our knowledge, it is the first work of connecting the pre-diagnosis information with machine learning-based multi-modal representation learning;
- According to similar pre-diagnosis patterns, we construct multiple graphs to represent the connection between patients. The graphs can regularize the patients' representations extracted from the diagnosis process and generate the optimized feature with a feature fusion mechanism.
- Extensive experiments on two datasets of mental disorders show that our way of introducing pre-diagnosis information can effectively improve diagnostic performance, and the introduction of the pre-diagnosis information conforms to intuition and medical logic.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 introduces our proposed multiple graph regularized diagnosis method. Section 4 gives extensive experimental results. Finally, Sect. 5 concludes the paper.

2 Related Work

2.1 Computer-Aided Diagnosis

In recent years, deep learning has been widely used in the medical field, most of which focus on medical imaging analysis. There are also some works on mental disorder diagnosis. For example, [11] based on computer vision analysis technology to predict attention deficit and hyperactivity disorder (ADHD) and autism

spectrum disorder (ASD). They considered the tester's facial expression, head position, movement, etc., and used a support vector machine (SVM) to analyze ADHD and ASD is classified. [9] build a multi-modal method using 3D facial expressions and spoken language with C-CNNs to predict PHQ scale. [18] process acceleration signal from wrist and ankle by RNN to distinguish behavior between ADHD diagnosed participants and normal ones. [2] uses clinical conversation messages to train a text-based multi-task BiLSTM network aiming at modeling both depression severity and binary health state. [17] try to diagnose the major depressive disorder or simply depression through electroencephalogram by Logistic regression, Support vector machine, and Naive Bayesian. [24] describe an intelligent auxiliary diagnosis System based on multimodal information fusion to diagnose ADHD. [26] use a cross-task approach that transfers attention from speech recognition to depression severity measurement. [3] use atrous residual temporal convolution network and temporal fusion based on visual behavior to capture depression signals. These works focus on the analysis in the diagnosis process, ignoring the integration of pre-diagnosis information, which is not enough to support a complete ADHD diagnostic logic chain.

2.2 Graph-Based Information Fusion

Constructing a graph to make feature fusion between related entities is widely applied in various applications. [20] build a multi-modal heterogeneous graph for recommendation. [23] design a heterogeneous graph neural network to jointly consider heterogeneous structural information and contents information of each node effectively. Using graph-based fusion model to simulate human logic has also been widely applied in the domain of computer-aided diagnosis. For example, [15] build the interaction between individuals and groups is established through a graph convolution network to assist in diagnosing COVID-19. [25] introduce a graph-based semi-supervised model by using partly labeled samples to diagnose dementia. [16] construct a knowledge graph to assist diagnosis based on doctor experience, case data, and other information. It proves the feasibility and research value of intelligent diagnosis assisted by graph network.

We combine the pre-diagnosis information with the diagnostic analysis process. First, the time-sequential neural network is to fuse the patient's various characteristics during the test. Then the graph neural network is established through the pre-diagnosis information, and the feature optimization is performed through the attention mechanism to improve the prediction effect.

3 Method

This paper proposed a MuGRED (Multiple Graph REgularized Diagnosis) model to solve the mental disorder diagnosis problem. The diagnosis process of the mental disorder can be summarized into two-step: Pre-diagnosis and diagnosis. During the pre-diagnosis process, patients are required to complete some questionnaires, and the doctors can get a preliminary understanding of their

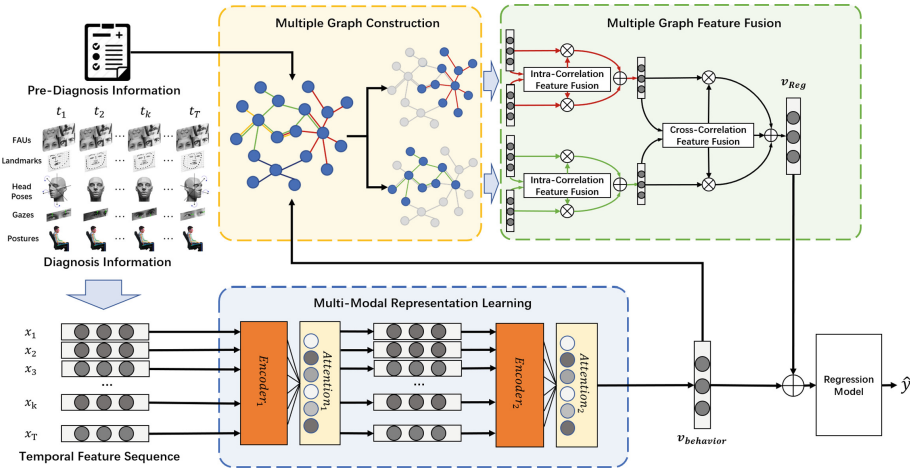

Fig. 1. Overview architecture of Multiple Graph REgularized Diagnosis (MuGRED).

psychological conditions based on the quantitative indicators. And during the diagnosis process, doctors interview patients or ask them to complete specific tasks and observe their multi-modal behaviors, such as eye movements, facial expressions, head posture, body movements, conversation content, or voice intonation. The doctors consider the patient condition of both pre-diagnosis information and performance of multi-modal behavior during the diagnosis process to draw the final diagnosis conclusion.

The proposed MuGRED model imitates the above diagnosis process of human doctors. As Fig. 1 shows, the MuGRED model consists of three modules: *Multimodal Representation Learning* module, *Multiple Graph Construction* modules, and *Multiple Graph Feature Fusion* modules. *Multimodal Representation Learning* module is to extract the behavior representation of the diagnosis process from the frame-level temporal multi-modal feature sequence. *Multiple Graph Construction* modules introduce the pre-diagnosis information to construct the correlation graph. The edges between patients are connected according to a similar pattern of pre-diagnosis results. *Multiple Graph Feature Fusion* modals firstly make a *intra-correlation feature fusion* to regularize the representation of patient behavior and then make a *cross-correlation feature fusion* to integrate the features regularized with different pre-diagnosis indicators for the final diagnosis conclusion making. In the following, we introduce the technical details of the MuGRED model.

3.1 Multi-modal Representation Learning

The inputs of the MuGRED model are from the diagnostic process. The patient behaviors are recorded as multimedia data, and we extract the patient's behavioral features according to the diagnostic logic needs, such as gaze position, facial

Action Units (FAUs) [8], facial landmarks, head pose, etc. The multi-modal behavior of the entire process is represented as a temporal feature sequence $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T]^T$ with T frames, where $\mathbf{x}_i$ is the frame-level feature of the timestamp i.

We need to extract the person-level feature to describe patient behavior from the temporal feature sequence. First, we split the sequence into N fragments with the same number of frames (Note that the overlaps between fragments are allowed), and the temporal behavior feature sequence can be transformed into a set of fragment sequences: $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N\}$. For each fragment sequence $\mathbf{c}_i$, we summarize and incorporate the contextual information of a frame with a temporal neural network, and transform the sequence of frame-level features $\mathbf{c}_i$ into a hidden state sequence $\mathbf{h}_i$, noted as:

$$\mathbf{h}_i = \mathbf{\Phi}(\mathbf{c}_i, \Theta) \tag{1}$$

where $\mathbf{\Phi}$ can be any temporal encoder, such as *Temporal Convolutional Network* (TCN) [14] or *Long-Short Term Memory Network* (LSTM) [10]. Then we aggregate the hidden state of each frame with a self-attention mechanism to generate the representation of fragment feature $\mathbf{r}_i$ as:

$$\mathbf{r}_i = \frac{1}{M} \sum_{i=1}^{M} \mathbf{w}_i \mathbf{h}_i \tag{2}$$

where $\mathbf{w}_i = \frac{e^{\mathbf{h}_i}}{\sum_{i=1}^{M} e^{\mathbf{h}_i}}$ is the attention weight and M is the frame number in fragment sequences $\mathbf{c}_i$. So that we obtain the sequence of fragments $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N]$. By repeating the same process of transforming the frame-level feature set $\mathbf{c}_i$ to the fragment feature $\mathbf{r}_i$, we can transform the fragment-level feature set $\mathbf{r}$ into the person-level representation $\mathbf{v}_{Behavior}$.

3.2 Multiple Graph Construction

After generating the person-level feature, we introduce the pre-diagnosis information. Given P kinds of pre-diagnosis patterns, the pre-diagnosis result of the patient i is represented as $\mathbf{R}_i = \{\mathbf{R}_i^1, \mathbf{R}_i^2, \cdots, \mathbf{R}_i^P\}$.

Suppose we've known the behavioral representation of $\hat{N}$ previous patients, we build the correlation hypergraph $G(\hat{V}, E)$ to represent the connections over the patients. Where the set of node $\hat{\mathbf{V}} = \{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \cdots \hat{\mathbf{v}}_{\hat{N}}\}$ denotes the person-level feature of the previous cases. As there are P kinds of pre-diagnosis patterns, the graph includes P kinds of edges, denoted as $\mathbf{E} = \{E^1, E^2, \cdots, E^P\}$. The *p-th* kind of correlation between node $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{v}}_j$ is denoted as:

$$e_{ij}^p = \begin{cases} 0 & \mathbf{R}_i^p \neq \mathbf{R}_j^p \\ 1 & \mathbf{R}_i^p = \mathbf{R}_j^p. \end{cases} \tag{3}$$

where $e_{ij}^p \in E^p$, $i, j = 1, 2, \cdots, \hat{N}$. i.e., if the node $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{v}}_j$ are with the same result of the *p-th* pre-diagnosis indicator, we defined that they are with

correlation in the corresponding aspect. Consequently, they are connected with the *p-th* kind of edge in the graph. For example, if two people are both with severe insomnia, we construct their correlation in the insomnia aspect. Note that the pair of nodes may contain multiple correlations, which indicates that they are with stronger correlations.

3.3 Multiple Graph Feature Fusion

In the previous section, we used the pre-diagnosis information to construct an association graph between existing cases $G(\hat{V}, E)$. Next, we obtain the regularized patient behavior feature with a two-step operation: First, for each correlation constructed by the specific pre-diagnosis pattern, we make a *intra-correlation feature fusion* to regularize the behavior feature; and then we make a *cross-correlation feature fusion* to aggregate the set of features regularized with the single correlation.

Intra-correlation Feature Fusion. As we obtained the personal-level behavioral feature vector $\mathbf{v}_{Behavior}$ of the current patient, we regularize the feature representation with the correlation graph. By adding the node $\mathbf{v}_{Behavior}$ into the correlation graph, we connect the edges along with the above rules based on pre-diagnosis results' consistency.

For the *p-th* kind of pre-diagnosis pattern, the feature $\mathbf{v}_{Beh}$ is optimized on the sub-graph of G, denoted with $g_p = (\mathbf{V}, \mathbf{E}^p)$. We aggregate the correlated feature representation with the current case by a *graph attention network* on each pre-diagnosis indicator. The correlation and message passing that the node $\hat{\mathbf{v}}_i$ affected the node $\mathbf{v}_{Beh}$ is considered as a multi-head attention mechanism, expressed as:

$$\begin{aligned} \xi^p_{k,i} &= LeakyReLU((\mathbf{W}^k_{self}\mathbf{v}_{Beh}) \cdot (\mathbf{W}^k_{obj}\hat{\mathbf{v}}_i)^T),\ \hat{e}^p_i \neq 0 \\ \alpha^p_{k,i} &= Softmax_j(\xi^p_{k,i}/s_{scale}) \\ \mathbf{s}^p_i &= \Big\Vert_{k=1}^{K} \sigma(\Sigma_{j\in E_p}\alpha^p_{k,i}\mathbf{W}^k_{fea}v_i)\mathbf{W}^O \end{aligned} \tag{4}$$

where K is the number of attention headers, $\mathbf{W}^k_{self}$, $\mathbf{W}^k_{obj}$ and $\mathbf{W}^k_{fea}$ are the linear transformation parameters of the k-th attention header. $\|$ represents concatenation operation, and the hyper-parameter s_{scale} is a scaled index to adjust the sensitivity of feature differences in practical applications. Finally, a dimensional attention matrix $\mathbf{W}^o$ is applied for dimensional reduction.

According to the above setting, we construct correlation graph of the previous patients through P kinds of pre-diagnosis indicators and obtain an optimized representation of the current case through each indicator. So that we get the regularized feature set $R_S = \{\mathbf{r}^1_s, \mathbf{r}^2_s, \cdots, \mathbf{r}^P_s\}$.

Cross-correlation Feature Fusion. The regularized features through various correlations of pre-diagnosis indicators describe the patient's characteristics from different aspects. To generate the pre-diagnosis-related feature of the patient, we

propose a cross-correlation feature fusion with attention mechanism to fuse the features in S. The attention weight of each feature is shown as:

$$\mathbf{w}^p = \frac{1}{P} \sum_{p \in P} q^T \cdot \tanh(\mathbf{W} \cdot s^p + b) \tag{5}$$

$$\beta^p = \frac{\exp(\mathbf{w}_i^p)}{\sum_{p=1}^{P} exp(\mathbf{w}^p)} \tag{6}$$

where $\mathbf{W}$ is the weight matrix, q is the attention vector. β^p is the normalized weight. And the final feature is $\mathbf{v}_{Reg}$:

$$\mathbf{v}_{Reg} = \sum_{p=1}^{P} \beta^p \cdot s^p \tag{7}$$

Finally, we make a concatenation to the feature directly extracted from the temporal behavioral feature sequence and the optimized feature to consider both patient's direct behavior and the correlations with the pre-diagnosis-related previous diagnostic experience. So the final feature representation of making the diagnostic decisions is denoted as $\mathbf{z} = [\mathbf{v}_{Behavior} \parallel \mathbf{v}_{Reg}]$ for the diagnosis result prediction.

4 Experiment

4.1 Dataset Description

We evaluate the performance of our proposed method on two datasets of mental disorders: A well-recognized mental disorder benchmark *DAIC-WOZ* and a practical on attention deficit and hyperactivity disorder *ADHD*, which are described as follows:

ADHD: We cooperate with The Children's Hospital of Zhejiang University School of Medicine to collect a dataset of patients with *Attention Deficit and Hyperactive Disorder* (ADHD), including 109 children for training and 15 for testing. Before the diagnosis begins, each patient and their parents must complete various psychological assessment scales during the pre-diagnosis process. With the patients' informed consent, we record videos of these children's behaviors during their diagnosis process with multiple cameras and extract their gaze, head poses, facial action units, facial landmarks, and body movements from the videos of the diagnostic process as a sequential temporal behavioral feature. In this work, we introduce the *Conners Comprehensive Behaviour Rating Scale* (*Conners*) [6] as the pre-diagnosis information. It is a questionnaire to gain a

better understanding of academic, behavioral, and social issues. The diagnostic conclusion is from the result of *SNAP-IV Teacher and Parent Rating Scale* (*SNAP-IV*) [21], which is an assessment of ADHD risk with score ranges from 0 to 3 from three main perspectives: *Inattenton* (Inatt for short), *Hyperactivity-Impulsity*(H/Imp for short) and *Oppositional Defiant Disorder* (ODD for short).

DAIC-WOZ: A well-recognized mental disorder benchmark that contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety [7], depression, and post-traumatic stress disorder. The participants are required to communicate with an animated virtual interview controlled by a human interviewer in another room, and their behaviors are collected as multimedia records, including video, audio and text collected. In 189 participants, 107 participants are used as training set, 35 as development set and 47 as test set. The diagnostic conclusion of DAIC-WOZ is evaluated with the depression index score Patient Health Questionnaire (PHQ-8) [13]. The questionnaire estimates the mental disorder with eight indicators: NoInterest, Depressed, Sleep, Tired, Appetite, Failure, Concentrating, and Moving. And the value of each indicator score ranges from 0 to 3 according to the severity, so the total score ranges from 0 to 24. In the experiment, we consider the multimodal behavioral features of FAUs, head poses, gazes, and 2D facial landmarks extracted from the interview videos as the input of the temporal behavioral feature sequence. We regard *gender* and each indicator in the *Patient Health Questionnaire* (PHQ-8) [13] scale as the patients' pre-diagnosis information. Since there is a strong correlation between the label and the indicators, introducing all the indicators as pre-diagnosis information is meaningless. Instead, we regard gender and randomly select several indicators as pre-diagnosis knowledge.

4.2 Implementation Details

To simplify the training process, we apply a transfer learning strategy. We first train the temporal feature extraction model to predict the label directly with 100 epochs and then transfer it to the *Temporal Feature Extraction* module of MuGRED with fixed gradient propagation to train the rest part of the model. We evaluate the diagnosis conclusion with the evaluation metrics of *Mean Absolute Error* (MAE) and *Root Mean Square Error* (RMSE).

The hyper-parameters for all the experiments are under the same setting. The model was optimized with the Adam optimizer to optimize the parameters, and we set the λ_1 to 0.9 and λ_2 to 0.999 with weight decay of 1e–4. The initialized learning rate is set to 1e–3. The final result for each experiment follows early-stop strategy and the number of train epochs is 500.

4.3 Baselines

We considered three baseline methods for multi-modal representation learning. *DepArt-Net* is one of the state-of-the-art method of on the *DAIC-WOZ* dataset problem, which applied an *atrous temporal convolutional network* to extract the multi-model sentence embedding from the multi-modal feature sequence, and then make the diagnostic decision [3], In addition, *HAN+TCN* and *HAN+LSTM* is the baselines of hierarchical attention network with two sequential encoders.

Table 1. The RMSE and MAE scores comparison of the different pre-diagnosis information aggregation methods on *DAIC-WOZ* dataset and *ADHD* dataset

Method	Concat	MuGRED	ADHD		DAIC-WOZ	
			MAE	RMSE	MAE	RMSE
DepArt-Net	×	×	0.507	0.613	4.665	6.004
	✓	×	0.424	0.568	4.626	5.892
	×	✓	0.403	0.470	2.780	3.914
	✓	✓	**0.400**	**0.465**	**2.461**	**3.171**
HAN-TCN	×	×	0.446	0.599	4.471	5.799
	✓	×	0.413	0.558	2.617	3.435
	×	✓	**0.352**	**0.461**	2.801	3.925
	✓	✓	0.363	0.472	**2.131**	**3.021**
HAN-LSTM	×	×	0.472	0.604	5.166	6.216
	✓	×	0.446	0.556	2.667	3.367
	×	✓	0.372	0.480	2.800	3.907
	✓	✓	**0.361**	**0.462**	**1.922**	**2.571**

4.4 Results

Considering that MuGRED introduced extra information of pre-diagnosis, to keep the comparison fair, we proposed another way to introduce the same information by simply concatenating the vector of pre-diagnosis indicators to the multi-model behavior embedding. We select gender and the indicator *NoInterest* as the pre-diagnosis information. The results in Table 1 compare the performance of the two kinds of pre-diagnosis information integration methods on the three kinds of temporal feature extractors. We found that for both datasets, the integration of pre-diagnosis can improve the diagnostic performance for the three kinds of feature extractors, which proves the value of introducing pre-diagnosis information. In comparing the two integration mechanisms, MuGRED leads to a more significant improvement, indicating that the model design of MuGRED plays the effectiveness of pre-diagnosis information. In addition, in most cases, integrating the pre-diagnosis information with both methods can make the best performance, which may become a practical trick of MuGRED's modification.

We then evaluate the impact of the amount of combined pre-diagnostic indicators on the diagnostic prediction performance on DAIC-WOZ dataset. As shown in Table 2, we compared the model performance in the three settings: without pre-diagnosis, with gender and 1 indicator (*NoInterest*), and with gender and 3 indicators (*NoInterest*, *Depressed* and *Sleep*). The results show that the richer the pre-diagnosis information introduced, the more accurate the prediction of the diagnosis conclusion.

Table 2. The RMSE and MAE scores comparison of the diagnosis results with different amounts of the pre-diagnostic indicators on *DAIC-WOZ* dataset

Method	w/o Pre-diagnosis		Gender+1 Indicator		Gender+3 Indicators	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
DepArt-Net	4.665	6.004	2.780	3.914	**2.378**	**2.876**
HAN+TCN	4.471	5.799	2.801	3.925	**2.294**	**2.871**
HAN+LSTM	5.166	6.216	2.800	3.907	**2.182**	**2.855**

Table 3. The RMSE and MAE scores comparison of the diagnosis items with/without pre-diagnostic indicators on *ADHD* dataset

Method	Inatt		H/Imp		ODD	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Without pre-diagnosis						
DepArt-Net	0.507	0.591	0.648	0.758	0.366	0.454
HAN-TCN	0.425	0.562	0.665	0.802	0.277	0.342
HAN-LSTM	0.500	0.554	0.647	0.770	0.402	0.443
With pre-diagnosis of conners indicators						
DepArt-Net	0.414	0.480	0.496	0.552	0.297	0.356
HAN-TCN	0.391	0.469	0.451	0.552	0.272	0.338
HAN-LSTM	0.342	0.447	0.432	0.526	0.350	0.464

For the ADHD dataset, the ADHD risk rating is consist of three items, i.e., *Inattenton* (Inatt), *Hyperactivity-Impulsity* (H/Imp) and *Oppositional Defiant Disorder* (ODD). We wonder whether the pre-diagnosis information correlation can benefit all the performance of the predicted diagnosis items. To achieve this, we evaluate how the pre-diagnosis affects the single item of the diagnosis result. The results of comparing the prediction results of the specific three symptom index with/without pre-diagnosis information are shown in Table 3. It could be found that the performance of every single item using the same method has been improved with the help of pre-diagnosis in most cases.

5 Conclusion

In this paper, we focus on the problem of mental disorders diagnosis, where both pre-diagnosis and diagnosis process information are important for medical experts to make a diagnosis. However, most existing methods in CAD only focus on the diagnosis process information while ignoring the pre-diagnosis information. Moreover, the patients are commonly treated as independent samples in previous CAD methods, while medical experts, in real applications, always need to connect patients for comparison. Therefore, we adopted the pre-diagnosis information to connect the patients and proposed a Multiple Graph REgularized Diagnosis (MuGRED) method for mental disorders diagnosis. The proposed MuGRED method is an end-to-end learning algorithm by fusing the representation from both pre-diagnosis and diagnosis processes for final diagnosis prediction. Extensive experiments on both ASD and ADHD demonstrated the effectiveness of the proposed MuGRED method for mental disorder diagnosis.

Acknowledgement. This work was supported in part by Program of Zhejiang Province Science and Technology (2022C01044), National Natural Science Foundation of China (U20A20387), the Fundamental Research Funds for the Central Universities (226-2022-00142, 226-2022-00051), Project by Shanghai AI Laboratory (P22KS00111), and the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (SN-ZJU-SIAS-0010).

References

1. Bandini, A., Green, J.R., Taati, B., Orlandi, S., Zinman, L., Yunusova, Y.: Automatic detection of amyotrophic lateral sclerosis (ALS) from video-based analysis of facial movements: speech and non-speech tasks. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 150–157. IEEE (2018)
2. Dinkel, H., Wu, M., Yu, K.: Text-based depression detection on sparse data. arXiv e-prints pp. arXiv-1904 (2019)
3. Du, Z., Li, W., Huang, D., Wang, Y.: Encoding visual behaviors with attentive temporal convolution for depression prediction. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), pp. 1–7. IEEE (2019)
4. Feng, F., Wu, Y., Wu, Y., Nie, G., Ni, R.: The effect of artificial neural network model combined with six tumor markers in auxiliary diagnosis of lung cancer. J. Med. Syst. **36**(5), 2973–2980 (2012)
5. Giger, M.L.: Computer-aided diagnosis of breast lesions in medical images. Comput. Sci. Eng. **2**(5), 39–45 (2000)
6. Goyette, C.H., Conners, C.K., Ulrich, R.F.: Normative data on revised conners parent and teacher rating scales. J. Abnorm. Child Psychol. **6**(2), 221–236 (1978)
7. Gratch, J., et al.: The distress analysis interview corpus of human and computer interviews. In: LREC, pp. 3123–3128 (2014)
8. Hamm, J., Kohler, C.G., Gur, R.C., Verma, R.: Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. J. Neurosci. Methods **200**(2), 237–256 (2011)

9. Haque, A., Guo, M., Miner, A.S., Fei-Fei, L.: Measuring depression symptom severity from spoken language and 3d facial expressions. arXiv preprint arXiv:1811.08592 (2018)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
11. Jaiswal, S., Valstar, M.F., Gillott, A., Daley, D.: Automatic detection of ADHD and ASD from expressive behaviour in RGBD data. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 762–769. IEEE (2017)
12. Kaushal, C., Bhat, S., Koundal, D., Singla, A.: Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images. IRBM **40**(4), 211–227 (2019)
13. Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B., Berry, J.T., Mokdad, A.H.: The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. **114**(1–3), 163–173 (2009)
14. Lea, C., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks: a unified approach to action segmentation. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9915, pp. 47–54. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_7
15. Liang, X., Zhang, Y., Wang, J., Ye, Q., Liu, Y., Tong, J.: Diagnosis of Covid-19 pneumonia based on graph convolutional network. Front. Med. **7** (2020)
16. Liu, P., et al.: HKDP: a hybrid knowledge graph based pediatric disease prediction system. In: Xing, C., Zhang, Y., Liang, Y. (eds.) ICSH 2016. LNCS, vol. 10219, pp. 78–90. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59858-1_8
17. Mumtaz, W., Xia, L., Ali, S.S.A., Yasin, M.A.M., Hussain, M., Malik, A.S.: Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). Biomed. Signal Process. Control **31**, 108–115 (2017)
18. Muñoz-Organero, M., Powell, L., Heller, B., Harpin, V., Parker, J.: Using recurrent neural networks to compare movement patterns in ADHD and normally developing children based on acceleration signals from the wrist and ankle. Sensors **19**(13), 2935 (2019)
19. Stoitsis, J., Valavanis, I., Mougiakakou, S.G., Golemati, S., Nikita, A., Nikita, K.S.: Computer aided diagnosis based on medical image processing and artificial intelligence methods. Nucl. Instrum. Methods Phys. Res., Sect. A **569**(2), 591–595 (2006)
20. Sun, R., et al.: Multi-modal knowledge graphs for recommender systems. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1405–1414 (2020)
21. Swanson, J.M., et al.: Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and odd symptoms at the end of treatment. J. Am. Acad. Child Adoles. Psychiatry **40**(2), 168–179 (2001)
22. Vyas, K., et al.: Recognition of atypical behavior in autism diagnosis from video using pose estimation over time. In: 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6. IEEE (2019)
23. Zhang, C., Song, D., Huang, C., Swami, A., Chawla, N.V.: Heterogeneous graph neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 793–803 (2019)
24. Zhang, Y., Kong, M., Zhao, T., Hong, W., Zhu, Q., Wu, F.: ADHD intelligent auxiliary diagnosis system based on multimodal information fusion. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 4494–4496 (2020)

25. Zhao, M., Chan, R.H., Chow, T.W., Tang, P.: Compact graph based semi-supervised learning for medical diagnosis in Alzheimer's disease. IEEE Signal Process. Lett. **21**(10), 1192–1196 (2014)
26. Zhao, Z., Bao, Z., Zhang, Z., Cummins, N., Wang, H., Schuller, B.: Hierarchical attention transfer networks for depression assessment from speech. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7159–7163. IEEE (2020)