DEEP FACTOR REGRESSION FOR COMPUTER-AIDED ANALYSIS OF MAJOR DEPRESSIVE DISORDERS WITH STRUCTURAL MRI DATA

Erkun Yang\textsuperscript{1}, Lihong Wang\textsuperscript{2}, David Steffens\textsuperscript{2}, Guy Potter\textsuperscript{2,*}, Mingxia Liu\textsuperscript{1,*}

\textsuperscript{1}Department of Radiology and BRIC, UNC at Chapel Hill, Chapel Hill, NC
\textsuperscript{2}Departments of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC

ABSTRACT

Major depressive disorder (MDD) is a prevalent and debilitating psychiatric mood disorder that affects millions of people worldwide. Conventional methods for MDD severity diagnosis usually rely on neuropsychological assessments that are subjective and susceptible. Recently studies have shown that structural MRI (sMRI) can provide objective biomarkers for MDD severity diagnosis. However, current MRI-based methods generally rely on hand-crafted imaging features and cannot explicitly identify MDD-associated depression symptoms, thus failing to increase our understanding of clinical and cognitive staging of MDD. In this paper, we first employ five depression symptom factors to quantitatively measure MDD grade from different aspects. Then, we design an end-to-end deep factor regression network (DFRN) to predict these factors directly from 3D T1-weighted sMRI scans. To uncover the contributions of different brain regions, we generate attention maps to uncover the implicit attention of the learned DFRN models. Experimental results on 116 MDD subjects show that the predictions for all five factors are positively correlated with ground-truth values. Attention maps also highlight the most informative brain regions for each factor.

Index Terms—MDD, sMRI, attention map

1. INTRODUCTION

Major depressive disorder (MDD), commonly termed as depression, is a common disorder affecting many people and is projected to become the second most debilitating disease worldwide by 2030 [1]. Many efforts have been devoted to MDD severity diagnosis/prognosis, but the pathological and biological mechanisms of depression remain largely unclear. Improving diagnostic procedures is expected to lead to more interpretable and reliable results to facilitate early intervention or even prevention for at-risk MDD subjects.

Conventional diagnosis of MDD severity mainly relies on neuropsychological assessments. The most commonly used diagnostic criteria are defined in the Diagnostic and Statistical Manual for mental disorders V (DSM-V) [2] and Hamilton Rating Scale for Depression (HRSD) [3]. A previous study has shown that clinicians’ consistency in diagnosing MDD through DSM-V is only 0.25 [4], which may negatively affect the effectiveness of MDD diagnosis and even reduce the credibility of psychiatry. Besides, diagnosis based on neuropsychological scores may be subjective, due to missing patient’s family history or exaggerated reports on patient’s symptoms. Therefore, it is important to objectively evaluate MDD grade in a more comprehensive manner.

Recently, rapid advances in multimodal neuroimaging techniques, such as structural magnetic resonance imaging (sMRI), have provided non-invasive and objective tools to study brain structure. Several sMRI-based methods have been proposed for automated classification and prediction of MDD [5] via machine learning techniques, such as support vector machines (SVM), Gaussian process classifier (GPC), and linear discriminant analysis (LDA). These studies generally rely on hand-crafted features for representing sMR images. However, hand-crafted features may not be optimal for subsequent diagnosis, and may result in a diagnostic performance decrease. Thereby, it is desired to incorporate feature extraction and model training into a unified framework for end-to-end learning. To this end, following [6], we first employ five depression symptom factors to quantitatively measure MDD grading, including Anhedonia, Appetite, Sleep, Suicidality, and Anxiety. We then propose a deep factor regression network (DFRN) to jointly extract sMRI features and predict/regress these factor scores. The proposed model can be trained in an end-to-end manner, with the sMR image as the input and each of five factor scores as the output. To better understand the underlying mechanism and leverage the learned information further, we generate voxel-precise attention maps via DFRN for each factor. Experimental results on 116 subjects show that all the factors’ predictions are positively correlated with the ground-truth values. Also, the obtained attention maps highlight several informative regions in sMRIs for each of the five factors.

2. METHODOLOGY

2.1. Definition of Depression Symptom Factors

Following [6], we employ five depression symptom factors to quantitatively measure the MDD grade in a comprehen-
Anxiety
Suicidality
Sleep
Appetite
Anhedonia
fully connected
Average Pool
Max pool
3 3 3 conv, 128
3 3 3 conv, 256
3 3 3 conv, 512
3 3 3 conv, 64
7 7 7 conv, 64

Fig. 1: Architecture of the proposed Deep Factor Regression Network (DFRN) for the regression of five depression symptom factors.

To summarize, we conduct a common factor analysis using principal component analysis (PCA) with varimax rotation to produce five orthogonal factors. Due to the orthogonal property, these factors can reflect depression symptoms from different aspects, thus helping objectively assess the MDD grade. Factor interpretations are as follows: Anhedonia and Sadness (Factor name: Anhedonia); Appetite and Weight loss (Factor name: Appetite); Sleep Disturbance (Factor name: Sleep); Suicidality and Guilt (Factor name: Suicidality); and Anxiety and Tension (Factor name: Anxiety). Each patient has five quantitative factor scores to objectively assess the MDD grade. Besides, each subject is represented by a 3D T1-weighted MRI scan.

2.2. Deep Factor Regression Network (DFRN)

To predict factor scores from 3D sMRIs, we design a Deep Factor Regression Network (DFRN) with the whole-brain sMRI (size: 182 × 218 × 182) as input and a specific factor score as output. Fig. 1 illustrates the architecture of the proposed DFRN, which consists of one convolutional (Conv) layer, one max pooling layer, four residual blocks, one global average pooling (GAP) layer, and one fully connected (FC) layer. The first Conv layer uses 7 × 7 × 7 kernels, 1 × 2 × 2 stride, and 3 × 3 × 3 zero padding, followed by batch normalization (BN), and rectified linear unit (ReLU) activation. The channel number is 64. The max pooling layer uses 3 × 3 × 3 kernel size, 2 × 2 × 2 stride, and 1 × 1 × 1 zero padding. The output from the max pooling layer is then forwarded to four stacked residual blocks. All these four residual blocks contain two Conv layers with 3 × 3 kernel, 1 × 1 stride, and 1 × 1 zero padding. The BN and ReLU activation are applied to the first Conv in the residual block, while only BN is applied to the second Conv. A skip connection is applied to sum the input of the residual block and the output of the second Conv layer. The ReLU activation is then applied to the summed output. The channel numbers for the Conv layers in the four residual blocks are 64, 128, 256, and 512, respectively. On the feature maps yielded by the last residual block, a GAP operation is applied to form a feature vector, which is then fed to the last FC layer to regress each factor.

With the above DFRN, we can obtain the prediction value of \( x_i \) for the \( k \)-th factor via

\[
y^k_i = \phi(x_i, \theta_k), i \in \{1, 2, \cdots, n\},
\]

where \( \phi(\cdot, \theta_k) \) denotes the network for the \( k \)-th factor and \( \theta_k \) is the corresponding parameter set. For each factor, the network is optimized with the following objective function

\[
\min_{\theta_k} \mathcal{L}^k = \sum_{i=1}^{n} \left\| \hat{y}^k_i - y^k_i \right\|_2^2,
\]

where \( \hat{y}^k_i \) is the prediction of the \( i \)-th instance for the \( k \)-th factor. Since we have five factors, an intuitive idea is to design a multi-task learning framework and jointly optimize the five factors. However, in this work, joint training may adversely affect the performance of these tasks. The reason is that those five regression tasks could be unrelated due to the orthogonal property of five factors. To verify this hypothesis, we conducted experiments by training five tasks simultaneously. The experimental results show that the regression performance generated by joint training is much worse than that of separate training. Therefore, in this work, we train an individual DFRN model for each of five factors, respectively.

2.3. Generation of Factor Attention Map

The end-to-end learning makes our DFRN a black box. To understand the mechanism underlying the algorithm, we need to provide interpretable results. Since the overall objective is to regress five factors for assessing MDD grade, we generate factor attention maps (FAM) to highlight the most informative brain regions relevant to each factor.

Inspired by [7], we first train the DFRN for each factor via Eq. (2) and then keep the network parameters fixed. For each input sMRI, we backpropagate its gradients to the input and obtain a voxel-precise spatial FAM for the corresponding factor. Such FAMs are expected to uncover the implicit attention maps are determined by the model parameters that

\[
R^k = (f^k > 0) \cdot (R^{k+1} > 0) \cdot R^{k+1}.
\]

Compared with the vanilla backpropagation, Eq. (3) adds additional guidance from higher layers. This can prevent the flow of negative gradients, corresponding to neurons that decrease the activation of the final units we attempt to visualize.

For each factor \( k \) and an sMRI \( x_i \), we can generate a specific attention map \( a^k_i \). To leverage the information from all samples, we further average all the attention maps for each factor to obtain a more reliable attention map as follows

\[
a^k = \frac{1}{n} \sum_{i=1}^{n} a^k_i,
\]

where \( a^k \) is the final FAM for the factor \( k \). Note that our model accepts the whole-brain image as input, and the attention maps are determined by the model parameters that
3. EXPERIMENTS

3.1. Materials and Experimental Setup

All participants were enrolled in Neurocognitive Outcomes of Depression in the Elderly (NCOED) Study [8]. This study enrolled two groups: (1) participants with a diagnosis of MDD, and (2) normal controls with no history of depression or other major psychiatric illness, who served as a comparison group. According to the research question, only the depressed group with \( n = 116 \) MDD subjects with 3D T1-weighted sMRIs were involved in this work.

A 10-fold cross-validation strategy was used in the experiments. Specifically, we first randomly split all subjects into 10 groups. Then, one group was alternatively used as test set and the remaining 9 groups as training set. To avoid bias in random partition, the experiments were run 5 times independently. The proposed DFRN model was trained with Adam algorithm (batch size: 4, learning rate: 0.001). All the experiments are implemented with PyTorch under the environment of Python 3.7 on the Ubuntu 18.04 system with NVIDIA TITAN Xp GPU.

To quantitatively analyze the results of DFRN, we report the Pearson correlations between the predicted scores and real scores in Table 1, and also visualization the prediction scores and the real scores in Fig. 2. Results in Fig. 2 and Table 1 suggest that, our predicted scores are positively correlated with the ground-truth scores for each of five factors. This demonstrates that DFRN can reliably and automatically predict the five factor scores based on whole-brain sMRI scans. We can also see that the Pearson correlations for Anhedonia and Sleep are relatively higher (with small \( p \) values). This indicates that the relationship between these two factors and sMRI scans is strong when identifying MDD patients from normal controls.

To compare DFRN with traditional methods with hand-crafted features, we extract volumes of gray matter tissue inside 170 regions-of-interest (defined in AAL3) as features and regress five factors via the ridge regression (RR) model. The Pearson correlation of RR and the average performance for DFRN are illustrated in Fig. 3, which clearly suggests the superiority of the proposed DFRN method.

3.2. Results and Analysis

3.2.1. Results of Factor Prediction

To quantitatively analyze the results of DFRN, we report the Pearson correlations between the predicted scores and real scores in Table 1, and also visualization the prediction scores and the real scores in Fig. 2. Results in Fig. 2 and Table 1 suggest that, our predicted scores are positively correlated with the ground-truth scores for each of five factors. This demonstrates that DFRN can reliably and automatically predict the five factor scores based on whole-brain sMRI scans. We can also see that the Pearson correlations for Anhedonia and Sleep are relatively higher (with small \( p \) values). This indicates that the relationship between these two factors and sMRI scans is strong when identifying MDD patients from normal controls.

Table 1: Pearson correlation (\( p \)-value) between the predicted and real scores for different factors in five runs.

<table>
<thead>
<tr>
<th></th>
<th>Anhedonia</th>
<th>Appetite</th>
<th>Sleep</th>
<th>Suicidality</th>
<th>Anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run #1</td>
<td>0.108 (0.248)</td>
<td>0.081 (0.385)</td>
<td>0.234 (0.012)</td>
<td>0.066 (0.020)</td>
<td>0.082 (0.385)</td>
</tr>
<tr>
<td>Run #2</td>
<td>0.179 (0.055)</td>
<td>0.128 (0.172)</td>
<td>0.254 (0.006)</td>
<td>0.100 (0.284)</td>
<td>0.104 (0.269)</td>
</tr>
<tr>
<td>Run #3</td>
<td>0.121 (0.197)</td>
<td>0.149 (0.112)</td>
<td>0.171 (0.067)</td>
<td>0.086 (0.362)</td>
<td>0.152 (0.104)</td>
</tr>
<tr>
<td>Run #4</td>
<td>0.113 (0.225)</td>
<td>0.095 (0.312)</td>
<td>0.192 (0.039)</td>
<td>0.048 (0.613)</td>
<td>0.059 (0.529)</td>
</tr>
<tr>
<td>Run #5</td>
<td><strong>0.216 (0.020)</strong></td>
<td>0.067 (0.476)</td>
<td>0.103 (0.271)</td>
<td>0.026 (0.782)</td>
<td>0.020 (0.831)</td>
</tr>
</tbody>
</table>

3.2.2. Generated Factor Attention Maps (FAMs)

We generate FAMs to locate the most informative brain regions for each of the five factors via Eq. 4, with results shown in Fig. 4. This figure shows that the attention maps for the five factors are overall consistent with each other. Moreover, the most discriminative regions are located in corpus callosum, hippocampus, cerebellum, insular, caudate nucleus, and brain stem, which are consistent with previous studies [9]. This demonstrates the reliability of our DFRN for MDD analysis.
4. CONCLUSION

In this work, we design a deep learning framework called DFRN to model the relationship between structural MRIs and five depression symptom factors. Factor attention maps are also generated to identify MDD-related regions in sMRIs. Experimental results on 116 MDD subjects suggest that the predictions are positively correlated with ground-truth scores for the five factors, which indicates that DFRN can effectively uncover the deep embedded local morphological features from structural MRIs. Also, the discriminative regions identified by our method are located in corpus callosum, hippocampus, cerebellum, insular, caudate nucleus, and brain steam.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki and approved by the Duke University Institutional Review Board.

6. ACKNOWLEDGMENTS

This work was supported by NIH Grant (No. MH108560).

7. REFERENCES


