

Machine Learning Models for Depression Patient Classification Using fMRI: A Study

Sanskriti Gupta and Rekha Vig(✉)

The North Cap University, Gurugram, India
sanskritiguptal6@gmail.com, rekhavig@ncuindia.edu

Abstract. Depression is a crucial entity of clinical research that imposes challenges to clinicians regarding its diagnosis and correct treatment. Machine learning technologies, are being used successfully for diagnosis of depression, especially using fMRI (functional Magnetic Resonance Imaging). The brain fMRIs are T2* weighted images captured using Blood Oxidation Level Dependent (BOLD) signals and show the functional activity of the brain. BOLD signal maps the activated/functional areas of the brain according the level of oxygen transmitted. Machine Learning and fMRI when taken together provide a tool for clinicians to identify the depressed patients from healthy controls. This paper reviews methodologies of past studies for depression classification using fMRI and machine learning.

Keywords: Depression · fMRI · BOLD · Machine learning

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) has emerged as one of the powerful tools in neuro imaging techniques for studying human brain functions and dysfunctions. It is among the primary techniques for revealing the communication between spatially remote different brain regions and provides a promising way to study interaction between those regions that are engaged simultaneously in a task. fMRI bold signal [1] is temporally coherent between functionally coupled regions forming functional brain networks. To identify the activated brain portions in fMRI data during resting (task-free) state or task instances is crucial neuroscience for planning surgery and diagnosis of psychiatric disease such as Alzheimer, Epilepsy, Dyslexia, Schizophrenia, Depression etc.

Depressive disorder is among the leading causes of functional impairment worldwide. Appropriate diagnosis of depression is a prerequisite for successful medication and to preserve a good health. Tools like brain imaging techniques are very important to researchers for better understanding of brain mechanism implicate in depressive disorder. fMRI is among the popular imaging methods used for studying mental disorders. fMRI measures the variations of oxygenation in blood to form discrete time signals. The signals are known as BOLD signals. fMRI measured using T2* weighted echo-planer imaging techniques are based on bold signals and provide a direct measure of neural activity in the brain. When fMRI data is collected the subject can either be in

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11620, pp. 685–696, 2019.
https://doi.org/10.1007/978-3-030-24296-1_55

the resting state or performing task such as tapping finger, listening music, visualizing facial expression.

Many methods have been used for diagnosing diseases and for classification of normal and diseased subjects. As per the current knowledge, there are no standard automated tools available for disease recognition in the hospitals.

In recent year, owing to increasing success of machine learning methods in the area of image processing these methods are being employed in neuro disorder diagnosis or classification using fMRI. Recent researches using machine learning [2] applied to fMRI neuro-imaging data have been successful in identification of depressed individuals based on their brain activity. Feature selection techniques were developed to filter out a subset of relevant features with an aim of developing robust learning models. In general, these techniques are used as a previous step to classification algorithms for improving prediction accuracy. In addition, they also identify a set of meaningful features that best discriminate classes.

According to Grotegerd, when fMRI data was analyzed using machine learning algorithms, information from fMRI patterns can directly differentiate depressive disorder.

Although much advancement have been made in classifying depressed patients from the healthy ones using in fMRI data, research is still going on in this area to make the system more reliable. This paper presents a review on various techniques for classifying depressed patients. The rest of the paper presents this review in Sect. 2, some performance parameters in Sect. 3 and conclusion in Sect. 4.

2 Literature Survey

The two different approaches for extracting functional connectivity in resting state fMRI are:

1. Model based,
2. Data Driven.

Model based method is the traditional method which utilized seed based voxel method to compute the temporal correlation between time series (BOLD) of a specific brain region and that of other regions. This traditional voxel analysis is widely used for identifying functional brain networks as it is simple, sensitive, and easy to interpret. The networks so formed are however restricted by selection of the seed voxels. The Data Driven method, in contrast to Model based doesn't require assuming a model for exploring functional connectivity. For discovering reliable functionally connected networks (resting state), various methods have been suggested such as principle component analysis, independent component analysis (ICA) [3] and clustering approaches based methods.

Based on these approaches we classify our review into these two broad categories:

2.1 Model Based

Croddock et al. in 2009 [4] proposed a model based method which utilizes Support Vector Classification (SVC) for classification of features obtained by two selection algorithms (one filter method and one wrapper method). Although 95% accuracy was obtained when best feature selection was taken but it also showed an accuracy of 62.5% in the case of no feature selection.

Zeng et al. in 2012 [5] introduced a multivariant pattern analysis method to identify the major brain depression. There is a significant difference in network and region of brain in depressed patient compared to the healthy ones. This strategy was utilized in this method. The fMRI images of each subject after preprocessing were subjected to feature extraction from functional connectivity matrix. The discriminative power of the features was extracted in Eq. 1 using kendall tau rank correlation coefficient, which can be defined for function connectivity feature 'j' as:

$$\tau_j = \frac{n_c - n_d}{m \times n} \tag{1}$$

Where n_d and n_c are the number of discordant and concordant pairs respectively. $m \times n$ are the total number of sample pairs.

Sato et al. in 2015 [6] described a machine learning approach to detect major depression in fMRI data. This model based technique utilized a particular machine learning algorithm for classification, MLDA (Maximum Entropy Linear Discriminant Analysis) with feature selection of voxels having most discriminative features. The results also showed that there was no difference in the classifier decision between patients with no medication and the major depressed patients taking antidepressants.

The traditional Support Vector Machine Learning approach was replaced by Weighted Discriminative Dictionary learning by Zeng et al. in 2016, [7] which depend on the similarity between the dictionary atom and samples to make the representation model more discriminative. In this learning method, an objective function was computed and then the optimization strategy was followed into two sub groups: updating and fixing until all the weights were fixed.

In 2017, Wang et al. [8] proposed a Functional brain network (FBN) based approach for depression disorder classification. The functional connectivity matrix obtained by sparse low rank matrix was first converted into binary matrices by replacing all non-zeros connectivity to 1. Then from this FBN eight graph based features were extracted using Fisher score in Eq. 2. Fisher score for each feature was defined in paper as

$$FS = \frac{p_1(q_1 - q)^2 + p_2(q_2 - q)^2}{p_1\sigma_1^2 + p_2\sigma_2^2} \tag{2}$$

Where q is the mean value of features for all the samples, p_1, p_2 are the numbers of samples in the two classes, σ_1, σ_2 is the variance of two classes, and q_1, q_2 is the feature mean value of the two classes.

The larger the fisher score more were the discriminative features. The sparse low rank model and eight graph based features showed promising results for recognition of major depression disorder. The method so proposed was compared on the basis of features and classifier used (Naïve Bayes, k-measure weight, Linear Discriminant analysis and SVM). SVM with eight features was demonstrated as a best in terms of accuracy. The schematic diagram for the proposed method is shown in Fig. 1 below.

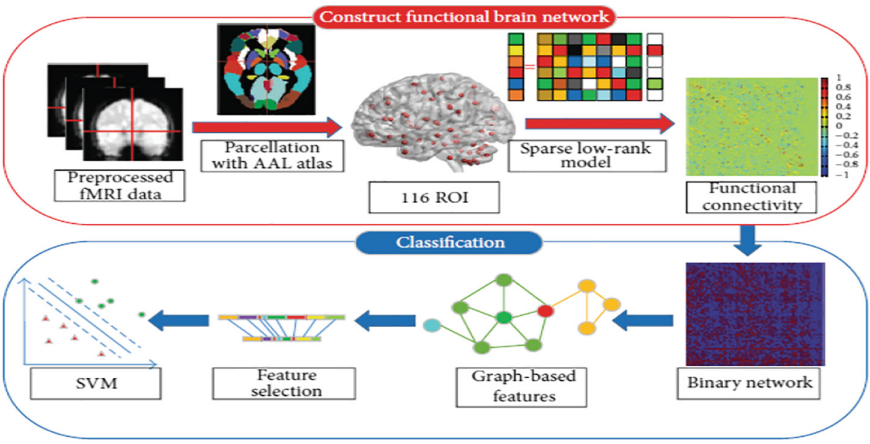

Fig. 1. Classification of depression disorder using functional brain network

2.2 Data Driven

A data driven method has been suggested by Sojoudi et al. in 2010 [9] where spectral clustering was utilized. An individual affinity matrix shows whether the brain region lies in one cluster or other. Network homogeneity as in Eq. 2 measures were used to evaluate the amount of similarity in a network which is derived as

$$Network\ Homogeneity = \frac{\sum_{i=1}^{n}(S_i)^2 - n(\bar{S})^2}{\frac{1}{12}K^2(n^3 - n)} \tag{3}$$

Where S_i is the sum rank of ith time point, $\bar{S}$ is the mean of S_i's, n is the total number of time points, and K is the number of regions within that network.

In 2012, Lu et al. [10] described data driven technique using event related fMRI data as well as resting- state data. On each data set PCA followed by ICA was applied and an absolute correlation coefficient d_{ij} in Eq. 4 was calculated as

$$d_{ij} = \frac{\sum_{t=1}^{T}\left(Se_i(t) - \overline{Se_l}\right)\left(Sr_j(t) - \overline{Sr_j}\right)}{\sqrt{\sum_{t=1}^{T}\left(Se_i(t) - \overline{Se_l}\right)^2 \sum_{t=1}^{T}\left(Sr_j(t) - \overline{Sr_j}\right)^2}} \tag{4}$$

Where Se_i and Sr_j are the event related and resting state components respectively. $i = (1, 2, 3 \ldots M), j = (1, 2, 3, \ldots N)$ Where N is the resting state independent components and M is the number of event related independent component. The spectrum obtained by this was analyzed for top event related component and first feature was selected which showed uneven energy distribution. Then by classification of groups with ROI analysis, the depressed patients were recognized from healthy ones.

Roudina et al. in 2014 [11] proposed a novel feature selection technique to identify the meaningful features that best discriminate the patient classes according to the following equation:

$$\hat{\beta} = \arg\min\ \beta \in \mathrm{R}^{\mathrm{p}} ||\mathrm{Y} - \mathrm{X}\beta||_2^2 + \lambda \sum_{\mathrm{k}=1}^{\mathrm{p}} |\beta_{\mathrm{k}}| \tag{5}$$

Where β is LASSO estimate, $\lambda \in \mathrm{R}^+$ is a regularization parameter, p is the number of features, X is data matrix and Y is label vector. On fMRI pre processed data, through LASSO sparse algorithm subset of the relevant features were created. The selection of features was accomplished after several iterations. The selected features which showed non zero coefficients after this whole procedure were subjected to SVM using nested leave one out cross validation scheme.

Jie et al. in 2015 [12] proposed a greedy search algorithm that moves in two directions forward as well as backward. During each iteration, forward selection that starts with empty set added a feature to the current set so as to reduce the loss function. Backward elimination, which starts from the complete set, removes a feature until a loss function exceeds a threshold. Both of these strategies combined with SVM known as SVM-FoBa is an adaptive greedy algorithm. When in forward step a new feature was added, backward step check the new feature subset. The loss function in Eq. 6 can be written as

$$\mathrm{J}(\mathrm{w}, \mathrm{b}) = \frac{1}{2}||\mathrm{w}||^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{n}} \left[1 - \mathrm{y_i}\left(\mathrm{w}^{\mathrm{T}}\mathrm{x_i} + \mathrm{b}\right)\right]_+ \tag{6}$$

Where $\mathrm{x_i}$ is the sample with $\mathrm{y_i}$ corresponding to its label, w os the weight vector, b is the offset, C is the hyperparameter and $[\cdot]_+$ denotes hinge function. This classification was not just limited to depression disorder from healthy control but it also discriminates bipolar depression from major depression. Hence two data set were used.

Rosa et al. in 2015 [13] describe a sparse network based model for classification of depressed patients and also showed some advantages over whole brain voxel-based approach. The sparse based model shows biologically useful network structures that best discriminates the subjects from controls. In addition to this, this approach was also compared with correlation based metrics models. Here two data sets were taken which were acquired from two different samples. In the first data set, sad faces of different emotions were shown to participants. In second data set, block related designs were

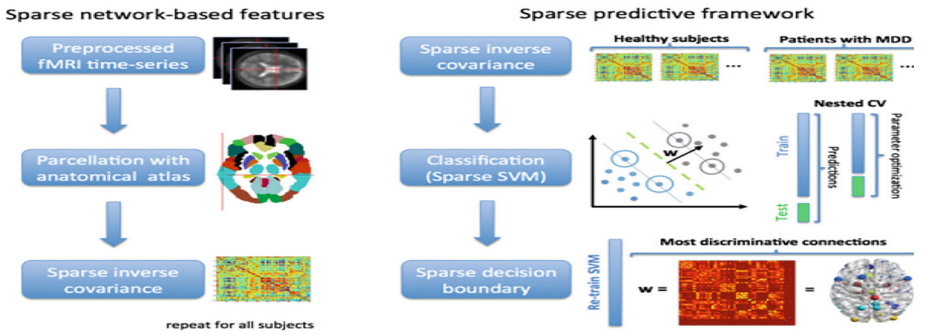

Fig. 2. Flow chart explaining the process of sparse network based model.

utilized where participants were allowed to view different emotions like happy, anxious, neutral and sad faces. The sparse network based proposed model is shown in Fig. 2.

A CCA and ICA based data driven method was proposed by Lin et al. in 2017 [14]. Here for detecting low frequency fluctuations a mixer model of both ICA and CCA was studied. After preprocessing of fMRI data, the cerebrum of each subject was divided into 90 regions using an Automated Anatomical Labeling (AAL). Fisher Discriminant Analysis was done prior to classification for the reduction of dimensionality of feature vector. Only 11 features were selected from these 90 regions. The results obtained from the mixer model were compared with filtered ICA, unfiltered ICA and CCA also.

A novel depression detecting algorithm was proposed by Miholca et al. [15] which utilized Relational Association Rules and Artificial Neural Network for machine learning. Functional connectivity was established based on the relation between different patterns of the activated brain regions. The RAR using fuzzy and crisp has been performed in this paper and the results were presented which shows almost accurate classification of healthy and depressed subjects. Feature selection plays a vital role during the classification.

3 Performance Parameters

To analyze the efficiency of various proposed machine learning approaches used for the classification of major depressive disorder patients from healthy controls, some parameters are studied. These parameters are accuracy, specificity, sensitivity. In regression based models, correlation coefficient is the measurement criteria.

Accuracy: It measures how accurately a test data classified. Mathematically it is defined as:

$$\text{Accuracy} = \frac{\text{PC} + \text{HC}}{\text{PC} + \text{PH} + \text{HC} + \text{HP}} \tag{7}$$

Sensitivity: It is the measurement to accurately identify the subjects. Mathematically it is defined as:

$$\text{Sesitivity} = \frac{\text{PC}}{\text{PC} + \text{PH}} \tag{8}$$

Specificity: It is the measurement to accurately identify the controls. Mathematically it is defined as:

$$\text{Specificity} = \frac{\text{HC}}{\text{HC} + \text{HP}} \tag{9}$$

Where PC, HC, PH, HP represent the number of patients classified as controls, healthy ones classified as controls, patients classified as healthy ones and healthy ones classified as patients respectively.

Correlation Coefficient: It measures how much the model prediction is correlated with associated label value.

Based on these parameters, the reviewed algorithms are compared and depicted in Tables 1 and 2.

Cross-validation is used to estimate the accuracy of an algorithm framed by specific machine learning method [16]. There are various methods for performing cross-validation.

1. Leave-one-out Cross-validation
2. Holdout Cross-validation
3. k-fold cross-validation: the sample.

The first two techniques i.e. Leave-one-out and Hold-out are the variants of k-fold cross-validation with little modification. K-fold cross validation technique splits the data into k sets of same size. After that, it works as follows: (1) classifies one set as test set from its k-set, and combine the other sets to form a training set, (2) training a model on training set by using the learning method, (3) the estimated model is tested on a test set, and (4) at last, computes validation parameter(s) to discover the model's reliability. This procedure is repeated for k-iterations, and at each time a new set classifying as the test set without repetition. The computed validation parameter(s) values are averaged from all the iterations to calculate the overall performance of the machine learning method.

Holdout approach is a k-fold cross-validation approach with k equals to 1. This approach is beneficial when the available data set has a large sample size. On the other hand, leave-one out cross- validation is beneficial when the available data set has considerably too small of a sample size. This approach is also similar to k-fold approach with k equal to the sample size of the data.

Table 1. Model based methods.

Author	Patient sample	Method	Description	Limitations	Result
Craddock et al. [4] (2009)	Subject-20 Control-20	Support Vector Classification (SVC)	Multiple Voxel Analysis using Support vector classification was employed as it produced good prediction accuracy and was less noise sensitive. Four feature selection approaches were compared	Too many iterations were used to remove the lower scoring features that could remove the important features and increase the computational expensive	Accuracy of 95% was achieved using 4 feature selection approaches. Even when no feature selection approach was employed the accuracy comes out to be 62.5%
Zeng et al. [5] (2011)	Subject-24 Control-29	Linear Support-Vector-Machine	To select the relevant features, Kendall tau rank correlation coefficient was utilized for whole brain resting state fMRI	AAL atlas covers the whole brain except the brain system, which can be the major study in depression Too many features can lead to over fitting	Accuracy-94.3% Sensitivity-100% Specificity-89.7%
Sato et al. [6] (2015)	Subject-25 Control-21	Maximum Entropy Linear Discriminate Analysis (MLDA)	A guilt selective change in functional connectivity was used for discrimination	Poor stability	Accuracy-78.26% Sensitivity-72% Specificity-85.71
Wang et al. [7] (2016)	Subject-29 Control-29	Weighted Discriminative Dictionary Learning (WDDL)	Separate dictionaries for patients and controls were formed and then classifier based on the low representation error was calculated for each new sample and hence the classification was done	Dictionary learning method treats all the samples indiscriminately and ignores the valuable relationship between the samples and dictionary atoms	Accuracy-79.31% Sensitivity-75.88% Specificity-82.76%

(*continued*)

Table 1. (*continued*)

Author	Patient sample	Method	Description	Limitations	Result
Wang et al. [8] (2017)	Subject-31 Control-29	Linear Support-Vector-Machine	Functional brain network was created using sparse low rank matrix. And from this functional connectivity matrix the features were extracted	Though it has remarkable accuracy but the construction of sparse low rank matrix was a complex one	Accuracy-95% Sensitivity-96.77% Specificity-93.10%

Table 2. Data driven method

Author	Patient sample	Method	Description	Limitations	Result
Sojoudi et al. [9] (2010)	Subject-16 Control-16	Spectral Clustering Technique	Individual Affinity matrix was constructed for each subject on which spectral clustering was done at group level. The clusters so obtained showed functional network presented in both healthy control and depressed patients	Clusters of data set may contain structures at different scales of density and size using Spectral Clustering	Network homogeneity was decreased in default mode network and visual cortex Network homogeneity was found to be consistent for both regions in patient group
Lu et al. [10] (2012)	Subject-13 Control-13	Pattern Recognition	The resting state as well as task based fMRI was combined for whole brain analysis after PCA and ICA and a correlation coefficient was used to compute the accuracy of the proposed lu algorithm	Whole brain analysis is difficult to interpret	Accuracy-81.82% Sensitivity-84.62% Specificity-77.78%
Roudina et al. [11] (2014)	Subject-30 Control-30	SVM linear kernel	Survival Count on random Subsamples feature selection approach was proposed which selects the stable features. The fMRI data was collected during visualization of happy faces	Limited to specific brain regions and ignores the useful information from whole brain analysis	Accuracy-72% Sensitivity-77% Specificity-67%

(*continued*)

Table 2. (*continued*)

Author	Patient sample	Method	Description	Limitations	Result
Jie et al. [12] (2015)	Subject-21 (BD), 25 (MDD) Control-20	Support Vector Machine-forward Backward elimination (SVM-FoBa)	In order to find the relevant features loss function was computed. In forward selection, the features were selected in order to minimize the loss function and in backward elimination, those features were selected by which the loss function threshold exceeds	Elimination of Voxel-wise analysis may discard the essential information for discrimination	Dataset 1 –Accuracy: 78.95% Sensitivity: 68.42% Specificity: 89.47% Dataset 2 –Accuracy: 85.00% Sensitivity: 83.33% Specificity: 86.67% Dataset 3 –Accuracy: 85.00% Sensitivity: 83.33% Specificity: 86.67%
Rosa et al. [13] (2015)	Subject-30 Control-30	–Sparse L1-norm support vector machines –Non-sparse L2-norm support vector machines	Two fMRI datasets of were taken, one was event based and another one was block based	Sparse models can be highly unstable in facilitating interpretations. Hence hinders their main aim	Dataset 1 –Accuracy: 78.95% Sensitivity: 68.42% Specificity: 89.47% Dataset 2 –Accuracy: 85.00% Sensitivity: 83.33% Specificity: 86.67%
Lin et al. [14] (2017)	Subject-20 Control-20	Linear Support-Vector-Machine (SVM)	Mixture model ICA and CCA was created and hence no need to filter data	Filtering may cause the loss of some useful information and the low frequency noises may ruin the results in the follow-up analysis	Accuracy-91.53% Sensitivity-90% Specificity-93.10%
Miholca et al. [15] (2017)	Subject-19 Control-20	RAR RBFN	fMRI dataset was collected during listening music	Low sample size. As with larger sample size, power of the classification model generalized well	Accuracy-100% Sensitivity-100% Specificity-100%

4 Conclusion

Depressive disorder is a common mental disorder which should be diagnosed in the early stages and accurately for better treatment. Various techniques based on machine learning have been developed so far and were discussed in this paper for the classification of depressed patients from healthy controls. These techniques were compared on the basis of accuracy, specificity and sensitivity. Support vector machine with leave one out cross validation and other machine learning algorithms were currently used and showed promising results. One more approach other than SVM is Rational Association Rules (RAR) and Artificial Neural Network (ANN) [15] which shows most accurate result with 100% accuracy. This paper showed some of the supervised as well as unsupervised machine learning approach for patient identification using fMRI and the review of these techniques was presented here.

Every technique has its own advantages and disadvantages. It is hard to state which technique is better as only accuracy is not sufficient for calculating the performance of a method, other parameters such as sample size is also a main consideration. As we have seen in the literature RAR shows 100% accuracy but the sample size is too small and may show poor accuracy with larger sample size that may not be acceptable.

References

1. Hongna, Z., et al.: The dynamic characteristics of the anterior cingulate cortex in resting-state fMRI of patients with depression. J. Affect. Disord. **227**, 391–397 (2017). https://doi.org/10.1016/j.jad.2017.11.026
2. Patel, M.J., Carmen, A., Julie, C.P., Kathryn, L.E., Reynolds III, C.F., Howard, J.A.: Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int. J. Geriatr. Psychiatry **30**(10), 1056–1067 (2015). https://doi.org/10.1002/gps.4262
3. Ilya, M.V., et al.: Whole brain resting-state analysis reveals decreased functional connectivity in major depression. Front. Syst. Neurosci. **4**(41), 1–10 (2010). https://doi.org/10.3389/fnsys.2010.00041
4. Craddock, R.C., Holtzheimer Ill, P.E., Hu, X.P., Mayberg, H.S.: Disease state prediction from resting state functional connectivity. Magn. Reson. Med. **62**(6), 1619–1628 (2009). https://doi.org/10.1002/mrm.22159
5. Zeng, L.L., et al.: Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. J. Neurol. **135**, 1498–1507 (2012). https://doi.org/10.1093/brain/aws059
6. Sato, J.R., Moll, J., Green, S., Deakin, J.F.W., Thomaz, C.E., Zahn, R.: Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. Neuroimage **233**, 289–291 (2015). https://doi.org/10.1016/j.pscychresns.2015.07.001
7. Wang, X., Ren, Y., Yang, Y., Zhang, W., Xiong, N.N.: A weighted discriminative dictionary learning method for depression disorder classification using fMRI data. In: Proceedings of the IEEE International Conference on Big Data and Cloud Computing and Networking, Sustainable Computing and Communication, pp. 618–622 (2016). https://doi.org/10.1109/bdcloud-socialcom-sustaincom2016.97

8. Wang, X., Ren, Y., Wensheng, Z.: Depression disorder classification of fMRI data using sparse low-rank functional brain network and graph-based features. Hindwai J. Comput. Math. Methods Med. 1–12 (2017). https://doi.org/10.1155/2017/3609821
9. Sojoudi, A., Seyed, M.S., Gholam, A.H.Z., Hamid, S.Z.: Spectral clustering of resting state fMRI reveals default mode network with specifically reduced network homogeneity in major depression. In: Proceedings of 17th Iranian Conference of Biomedical Engineering (2010). https://doi.org/10.1109/icbme.2010.5704952
10. Lu, Q., Liu, G., Zhao, J., Luo, G., Yao, Z.: Depression recognition using resting-state and event-related fMRI signals. Magn. Reson. Imaging **30**, 347–355 (2012). https://doi.org/10.1016/j.mri.2011.12.016
11. Rondina, J.M., et al.: SCoRS- a method based on stability for feature selection and mapping in neuroimaging. IEEE Trans. Med. Imaging **33**(1), 85–98 (2014). https://doi.org/10.1109/tmi.2013.2281398
12. Jie, N.F., et al.: Discriminating Bipolar Disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans. Auton. Ment. Dev. **7**(4), 320–330 (2015). https://doi.org/10.1109/tamd.2015.2440298
13. Maria, J.R., et al.: Sparse network-based models for patient classification using fMRI. Neuroimage **105**, 493–506 (2015). https://doi.org/10.1016/j.neuroimage.2014.11.021
14. Lin, W., Wu, H., Liu, Y., Lv, D., Yang, L.: A CCA and ICA-based mixture model for identifying major depression disorder. IEEE Trans. Med. Imaging **36**(3), 745–756 (2017). https://doi.org/10.1109/tmi.2016.2631001
15. Miholca, D.L., Adrian, O.: Detecting depression from fMRI using relational association rules and artificial neural networks. In: Proceedings of IEEE Conference on Intelligent Computer Communication and Processing, pp. 85–92 (2017). https://doi.org/10.1109/iccp.2017.8116987
16. Patel, M.J., Alexander, K., Howard, J.A.: Studying depression using imaging and machine learning methods. Neuroimage **10**, 115–123 (2016). https://doi.org/10.1016/j.nicl.2015.11.003