**PROCEDURE TO FOLLOW WHILE USING THE TOOLBOX**

Please cite the following paper [1] if you this toolbox useful for your needs.

**PREREQUISITES:**

* Please load the NIFTI toolbox, HRF-\_EST\_Toolbox, SPM8 toolbox in the Matlab path. I suggest you to put all toolboxes inside the 'toolbox' folder in the Matlab main folder, and add those paths permanently to Matlab startup.
* Have preprocessed data ready.
* Put all data (NIFTI files: nii files only) in the folder (nothing else should be in that folder).
* Make sure that the individual subjects for each disorder are placed in separate individual folders.
* Run Matlab as administrator on Windows only.

**Described here is the detailed procedure you may follow in order to use the toolbox most efficiently.**

1. Set the path to the data folder using the Browse button on the toolbox. After setting the Matlab Path, the path should appear in the Working Directory box.

Ex: working directory\Controls\Subject\_001\xxxxxx\nii.

Working directory\Controls\Subject\_002\xxxxxx\nii.

Working directory\PTSD\Subject\_001\xxxxxx\nii.

1. Choose the template for extracting the time series from the data, using the pop-up menu under the ROI Parcellation for time series option. CC200, CC400 and Dosenbach’s 160 are the templates implemented. These templates can be found in the Templates folder. If the user wants to use a different template, he can do so by using the browse option.
2. To calculate the Static FC, Static EC and Dynamic FC values for the Deconvolved data, select the Deconvolution checkbox and the required checkboxes from the ‘For-Deconvolved Data’ options. The deconvolution is performed according to the procedure detailed in Wu et al. [5].
3. To calculate the Static FC, Static EC and Dynamic FC values for the Non-Deconvolved data, select the required checkboxes from the ‘For Non-Deconvolved Data’ options.
4. The desired classifier required for the classification procedure can be chosen from the Classifiers list. The list of classifiers implemented in the toolbox is listed later in this document.
5. The user needs to specify the data split ratio value to divide the dataset into train and test datasets. The value inserted must be of decimal format.

Ex: If a value of 0.7 is entered, then the dataset is split into a train dataset which is 70% of the original dataset created or loaded and 30% is test dataset.

1. If the user doesn’t require to calculate the Static FC, Static EC and Dynamic FC values from their data and rather want to just use the classifiers, they need to select the load data checkbox and choose browse option to select the dataset they wish to classify and then the user can go on to choose the classifiers. (Default settings works best for this step, i.e. options required for the steps 1-4 be unselected.)

Ex: Table186 present in the folder is a sample dataset.

1. The toolbox has an option to further reduce the features by Recursive Cluster Elimination (RCE) [6] in the inner k-fold of the nested cross-validation procedure. The user can choose between the With-RCE and Without-RCE options, by selecting the respective radio button. If the user chooses the Without-RCE, they need not enter the values for the parameters and so, the parameter options disappears and if the user chooses the RCE option, he needs to enter the values for the given parameter options.
2. The Hyperparameter values need to be entered in the textboxes provided. The number of hyperparameter sets required varies with each classifier. So, select a classifier and then enter the lower limit value, the value by which you want to increment the lower limit value and the upper limit value.
3. The user can enter the desired values for the number of folds required for the k-fold cross-validation procedure, number of repetitions for the cross-validation procedure and number of classes. These values must be entered and cannot be left blank.
4. The user can choose the desired checkboxes placed on the bottom left of the toolbox based upon the user’s requirements.
5. The select significant checkbox performs a T-test/ANOVA to select the paths whose means are significantly different between the groups. This reduces the connectivity paths by orders of magnitude and speeds up the computation and reduces overfitting.
6. The hyperparameter search checkbox gives the user to optimize the hyperparameters for the models using a grid search. The optimization of the hyperparameters happens inside the inner k-fold loop in the classification procedure. This is done to obtain an unbiased estimate of classification performance.
7. The save decision surface checkbox gives the user the option of saving the decision surfaces obtained from the classification. This can be useful if you plan on using the saved decision surfaces/classifier models to classify unseen data. We recommend you to check this box on always.
8. The final option of max accuracy decision surface only saves the best decision surface obtained from the outer k-fold after either parameter optimization or Recursive Cluster Elimination (RCE) or both. This saves enormous amount of space on the hard drive as only the best classifier models for each partitioning (k-fold x repetitions) of the data. This option is only valid when save decision surface is checked and either hyperparameter search or RCE checkbox are checked as well.
9. Once the user has set the path, chosen all the required options, entered all the required values, they can click on ‘RUN’.
10. The desired results from the toolbox are saved in the ‘output’ folder which is created in the Working Directory that the user has set.

The following folders can be found in the ‘output’ folder:

* The deconvolution Folder contains the following folders:
  + run\_deconv – the nifti files that are converted to .mat files are saved in this folder.
  + Save\_deconv – The deconvolved files are saved in this folder.
  + Post\_decon\_extract – the extracted time series files post the deconvolution are saved in this folder.
  + Params
  + DFC, SFC, SEC- These folders contain the respective results and the excel file from each procedure chosen for deconvolved data.
* DFC, SFC, SEC- These folders contain the respective results and the excel file from each procedure chosen for the non-deconvolved data.
* Time-series Extract- the extracted time series files for the non-deconvolved data are saved in this folder.

14. A separate folder named ‘Rce-Results’ and ‘Outside\_Rce\_Results’ is also created in the Working Directory of the user. The ‘Rce-Results’ folder contains the classification results of ELM, KNN, LDA, LINEAR-SVM, NAIVEBAYES, QDA, RBF-SVM. The ‘Outside\_Rce\_Results’ folder contains the results of the remaining classifiers. This folder contains results from the classification process.

* The max\_Accuracy file which is obtained when the best accuracy for every partitioning of the data if hyperparameter optimization or RCE is checked.
* The Decision\_Surfaces folder which is obtained when the save decision surface option is checked. It saves the classifier models obtained by the classification procedure.
* Accuracy\_vals folder saves the accuracy values for every partition of the data (k-fold X no of repetitions)
* bestparameters file saves the best hyperparameters obtained from the inner-cross validation procedure.
* If RCE is selected or if the classifier provides us with feature importance scores then a mat file called feature\_impscore is present.
* The workspace of the classification procedure is saved as Results.

The classifiers implemented in the toolbox include

1. Probabilistic/Bayesian methods: Gaussian Naïve Bayes (GNB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Sparse Logistic Regression (SLR), Ridge Logistic Regression (RLR),
2. Kernel methods: Linear and Radial Basis Function (RBF) kernel Support Vector Machines (SVM), Relevance Vector Machines (RVM)
3. Artificial Neural Networks: MLP-Net (Multilayer Perceptron Neural Net), FC-Net (Fully Connected Neural Net), ELM (Extreme Learning Machines), LVQNET (Linear Vector Quantization Net)
4. Instance-based learning: K-Nearest Neighbors (KNN)
5. Decision Tree based Ensemble Methods: Bagged trees, Boosted Trees, Boosted Stumps, Random Forest, Rotation Forest.

More information about the classifiers can be obtained from Lanka et al. [2].

15. The Consensus Classifier option can be selected if the user wants to get the consensus accuracy among certain classifers. For this the user needs to select the Decision Surface Flag and the Maximum accuracy decision surface options before choosing the respective classifiers. Only the folders of the classifier results required for the consensus accuracy must be present in the ‘Rce\_Results’ folder and ‘Outside\_Rce\_Results’ folder.

# **References**

|  |  |
| --- | --- |
| [1] | L. Pradyumna, D. Rangaprakash, M. N. Dretsch, J. S. Katz, T. S. Denney Jr. and G. Deshpande, "Resting state functional connectivity data and a toolbox for automated disease diagnosis for Neurological disorders," *Data in Brief,* vol. Under Review, 2017. |
| [2] | G.-R. Wu, W. Liao, S. Stramaglia, J.-R. Ding, H. Chen and D. Marinazzo, "A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data," *Medical Image Analysis,* vol. 17, no. 3, pp. 365-374, 2013. |
| [3] | G. Deshpande, Z. Li, P. Santhanam, C. D. Coles, M. E. Lynch, S. Hamann and X. Hu, "Recursive Cluster Elimination Based Support Vector Machine for Disease State Prediction Using Resting State Functional and Effective Brain Connectivity," *PLoS ONE,* p. e14277, 2010. |
| [4] | P. Lanka, D. Rangaprakash, M. N. Dretsch, J. S. Katz, T. S. Denney Jr. and G. Deshpande, "Supervised Machine Learning for Neuroimaging-based Diagnostic Classification," *Neuroimage,* vol. Submitted, 2017. |