
Resting State Analysis: Simulation and validation of RS-fMRI dataset for ADHD Subjects

P. UDAYA PRASANTH^{a*}, O. UMA MAHESWARI^b

^aResearch Scholar, Department of ECE, Anna University, Chennai

^bProfessor, Department of ECE, Anna University, Chennai

*Corresponding author

Abstract

Novel methods for the analysis of functional magnetic resonance imaging (fMRI) data are being reported lately. It is necessary to validate these methods for reliability as the interpretations of the results are subjective as the ground truth in the fMRI data is not known. Validation of analysis methods requires knowledge of the ground truth of the data. Simulation studies are necessary to assess the quality of the statistical technique/analysis methods. The simulated fMRI dataset provided by various research institutions and researchers are mostly event/task-related. Resting-state fMRI analysis has been gaining importance recently for its ability to be used as a biomarker for various psychopathological conditions. Hence there is a need to generate simulated data for evaluation of the resting-state fMRI data analysis methods. In this paper, a method is proposed to simulate a complete 4D resting-state fMRI data using MATLAB. The fMRI data is simulated for normal and ADHD subject and the results are compared with real time data.

Keywords: *fMRI, resting state analysis, data simulation*

INTRODUCTION

The study of brain's functional anatomy has never been done so extensively as in fMRI analysis. fMRI is non-invasive, safe, painless and does not use harmful radiations. fMRI data acquisition is noisy and the analysis methods have evolved over time to give better understanding on functioning of the human brain. With new analysis methods being proposed every day, the interpretation of the results are questionable as the ground truth of the real fMRI data used is not known. These analysis methods need to be validated for reliability using simulated studies. There is no common data generating process available [1]. Various models have been proposed to simulate fMRI data by the researchers all over the globe[2]. Almost all data simulation methods are for event-related / task fMRI. Only a few simulators generate resting state fMRI dataset[3].

Resting-state fMRI (RS-fMRI) is the study of spontaneous brain activity. Spatio-temporal patterns of correlated activity in the absence of external stimulus were observed in the BOLD fMRI signals[4]. A set of brain regions exhibit strong low-frequency oscillations coherent during the resting state. These regions are collectively known as Default Mode Network[5]. The brain interrupts its intrinsic activity to respond to an external stimulus and do focused work. The brain is active all the time and a focused brain activity consumes only marginally (5%) more energy than at rest[6]. Task-based fMRI requires patient participation whereas RS-fMRI is easier to acquire, as it involves lesser patient involvement[7]. Recent studies on RS-fMRI suggest a valuable prognosis with good indicators/markers for understanding various psychopathological conditions. The Default Mode Network (DMN) of Attention Deficit Hyperactivity Disorder (ADHD) subjects show a distinct difference to typically developing(TD) individuals[8]. By selecting appropriate features, ADHD can be

differentiated from healthy subjects[9]. Complexity indicators in Down's Syndrome show higher complexity compared to the control group[10]. A consistent pattern of local abnormalities in the resting state of Autism Spectrum Disorder(ASD) subjects are potential neurobiological markers[11][12]. Artificial Neural Networks (ANN) have been used to classify ASD from normal subjects using the RS-fMRI dataset[13]. Functional connectivity analysis of RS-fMRI has been reported as a potential biomarker for Alzheimer's Disease[14]. Another resting-state analysis study suggests Functional connectivity (FC) between the olfactory network (ON) and hippocampus may be a sensitive marker for Alzheimer's disease (AD) progression[15]. Schizophrenia is another condition that can be studied using resting-state fMRI. Subjects with Schizophrenia show reduced network connectivity in self-referential networks and DMN[16], [17]. In subjects with Parkinson's Disease (PD), functional connectivity is reduced in cognitive networks, mainly in DMN, which could be a potential biomarker[18]. It is possible to monitor and stage PD using fMRI [19]. The list continues to grow and therefore it is imperative to evolve methods that study and analyze RS-fMRI as a possible prognostic tool. A reliable known data source is required for the validation of these methods. In this paper, a method is proposed to generate a complete 4D simulated resting-state fMRI dataset. The simulated dataset with known components is tested with Infomax ICA. The components are extracted and compared with real dataset. All the implementation is done within the MATLAB framework.

METHODOLOGY

A 4D fMRI dataset is a combination of Spatial Maps (SM) and its corresponding Time Courses (TC). To generate a simulated resting-state fMRI dataset, we need both spatial (SM) and temporal (TC) components. Spatial Maps are the activation regions where a corresponding time course is identified. The baseline spatial data for SMs are derived from an existing fMRI dataset[20]. The TCs of a resting state brain is characterized by very low-frequency fluctuation in the BOLD signal. The BOLD RS-fMRI has a frequency range of 0.01 to 0.25 Hz [21]. The Amplitude of Low-Frequency Fluctuations (ALFF) is a parameter representing the intensity of the spontaneous neural activity i.e., resting state[22]. With the known amplitude and frequency range the TCs are generated using the Inverse Fourier Transform[23]. For our purpose, time courses are modeled as band-limited (0.01 to 0.25Hz) random signals. Noise signals are added to the time courses and integrated into the spatial maps to complete the 4D dataset. An outline of the algorithm for the generation of a 4D fMRI dataset is given below

1. From an existing resting-state fMRI dataset, using single-subject data, create baseline spatial data by averaging the voxels over time points and thresholding the intensity to remove the non-brain voxels.
2. With a defined frequency band, using inverse Fourier transform, resting-state time course is generated
3. Noise signals – Physiological, System, and low drift noise are added to the time course[24].
4. Activation regions are selected and using the baseline spatial data, the time courses are embedded to get 4D volumetric data for resting-state fMRI.
5. Write the data back to NIFTI format for use with different analysis tools.

Following the above steps, a researcher can generate his resting-state fMRI dataset for validation of his analysis method. MATLAB with supporting toolbox for fMRI i.e., SPM12 and GIFT is used to implement the above steps.

The baseline volume data is derived from an available resting-state fMRI dataset. New York dataset from 1000 functional connectomes project is used in this work. The dataset is available in NIfTI format from the NITRC website[20]. The New York dataset has resting-state data from 84 healthy subjects and 25 ADHD subjects. All the subjects were evaluated psychiatrically. The details of the dataset are given below in table 1.

No. of healthy subjects	84 (43M / 41F)
Age (range)	7 - 49
No. of ADHD Subjects	25 (19M / 4F)
Age (range)	20-50
TR (Repetition Time)	2s
No. of slices	39
Time points	192
Image size	64 x 64

Table 1. Demographics of New York dataset

BASELINE SPATIAL DATA GENERATION

Baseline spatial data is obtained from a single subject data from the dataset. A normal subject data is chosen for the baseline volume. The baseline value of each voxel is the average of the measured time series of all the voxels[2]. Prior to averaging, The rest fMRI data is preprocessed using the SPM toolbox[25]. Preprocessing the fMRI data eliminates noises. The preprocessing steps used are

- i. Realignment
- ii. Normalization
- iii. Smoothing

The slices of each volume data are realigned over time points to eliminate the rotational and translational artifacts. All the slices need to be in the same orientation. The data is resliced to a voxel size of 3mm x 3mm x 3mm and spatially normalized using the standard MNI template. In group processing of fMRI data, normalization transforms each brain data to the same size, shape, and dimension. The results are then smoothed using a Gaussian kernel with full width half maximum (FWHM) value of 10mm to improve the signal-to-noise ratio. The value of FWHM was chosen based on the dataset. The smoothed volume data is averaged along the 192 time points. The volume size of the baseline data is 53 x 63 x 52 voxels. The non-brain voxel are made zero by thresholding the intensity value to 200. The mean voxel data to be used as baseline spatial data is obtained using the GIFT toolbox[26]. The orthographic view of preprocessed volume and baseline spatial data are shown in figure 1.

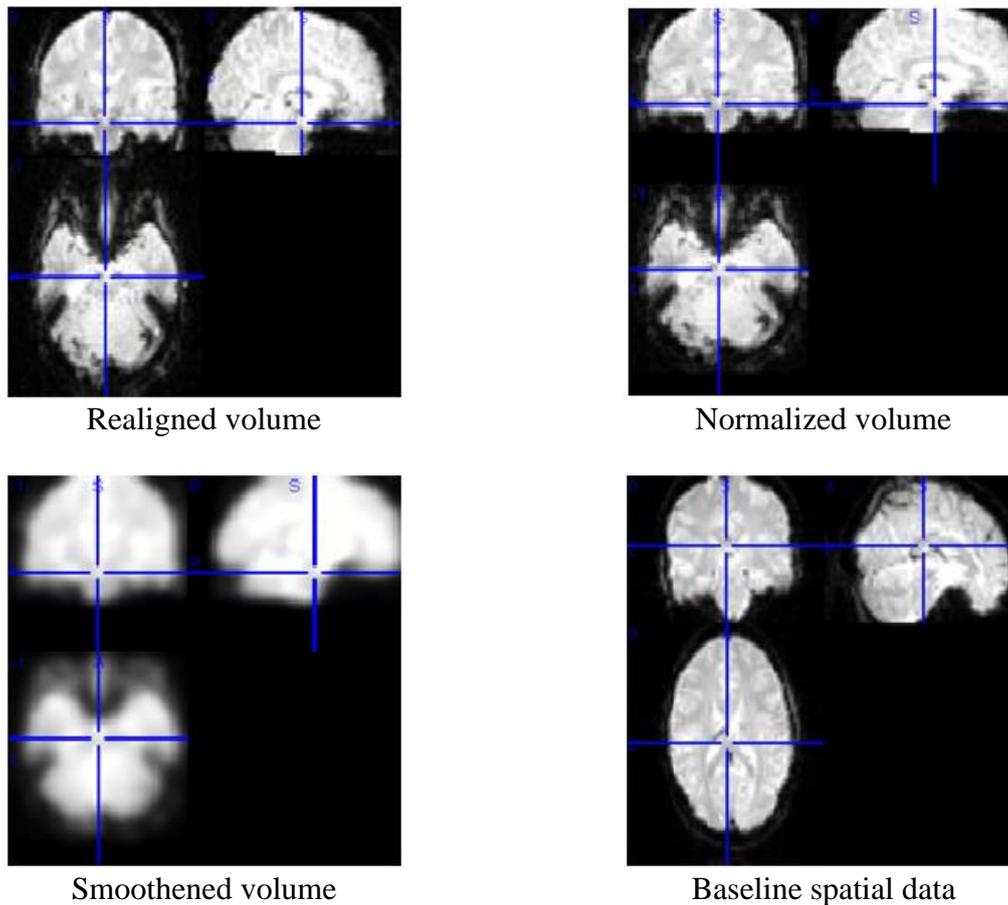


Figure 1. Orthographic view of the preprocessed and baseline spatial data

All the spatial noises have been minimized and baseline spatial data has been generated using an existing dataset.

TIME COURSE GENERATION

The resting-state time series is generated using the discrete inverse Fourier transform[23]. [23] describes a method for generating a random signal of known spectral form. A real, time-domain signal can be generated by defining the desired transform complex-frequency domain plane and inverse Fourier Transforming it. The resting-state fMRI has a very low-frequency band in the range of 0.01 – 0.25Hz[21]. In this work, we select a frequency range of 0.01 to 0.1 Hz. From the method described in [23], a power spectrum is constructed for the given band. The frequency band is normalized. The real part of the inverse Fourier transform is normalized and multiplied with the amplitude specified to generate the time series. The length of the time series is taken as 192 from the original dataset. In our work, we generate six time courses using the above method. The generated time series and its spectrum are shown in figure 2 and figure 3 respectively.

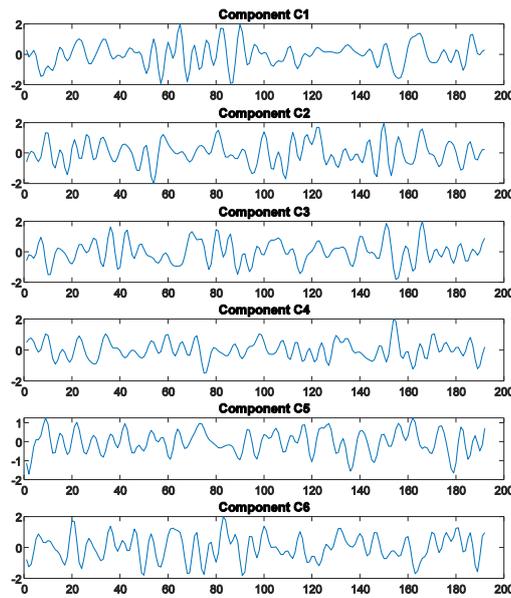


Figure 2. Generated resting state time series

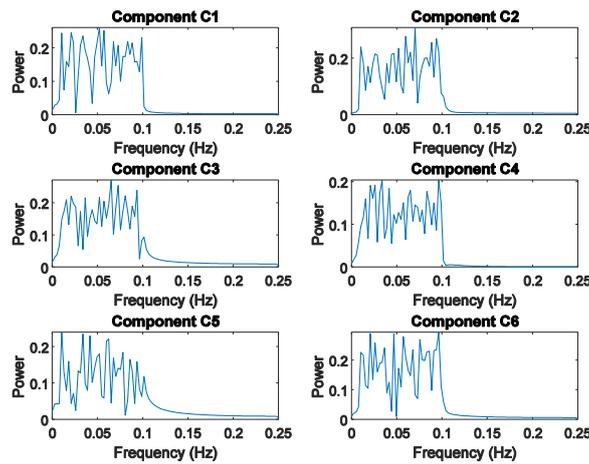


Figure 3. Power spectrum of generated components

The time series generated is then added with all the preferred noise and embedded into the baseline volume data.

NOISE ADDITION

The generated resting-state time series can be embedded directly into the spatial data, however, most real-time fMRI data are noisy and adding noise to our generated time series mimics the real data. The SNR of the noise signal is defined by the ratio of signal power to noise power. It can also be defined as the ratio of mean to standard deviation.

$$SNR = \frac{\mu}{\sigma}$$

Where, μ is the mean of the signal and σ is the standard deviation of the noise. Noise functions can be generated using the desired SNR. Noise in real fMRI data can be caused by various sources. We limit our discussion to system noise, physiological noise, low-frequency drift, and temporal noise. System noise is represented with a white Gaussian noise function. The physiological noise is caused by cardiac and respiratory disturbances. The noise is modeled as a function of sine and cosine function with user-defined frequencies. The default frequency of heartbeat and respiratory rate is chosen as 1.17 Hz and 0.2 Hz[2]. The frequency band of physiological noise overlaps with the fMRI band. The low-frequency drift is caused by the scanner system and can be considered system noise. It is modeled as a basis of discrete cosine functions[2]. The temporal noise is represented using autoregressive noise model $AR(p)$ of specified order. All the noise discussed is shown in figure 4.

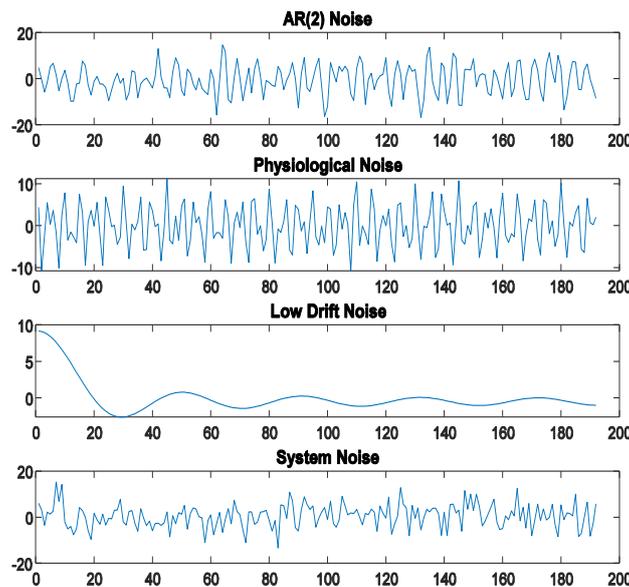


Figure 4. Noise types added to the time series

4D VOLUMETRIC FMRI DATASET GENERATION

To complete the generation of the 4D volumetric resting-state fMRI dataset, the generated time series needs to be embedded into the spatial data. The baseline volume data was derived from an existing dataset and preprocessed. Adding time courses to the 3D volume makes it a complete fMRI dataset. The time courses were simulated and the noise was added to it. The fMRI data analysis reveals that neural activity usually occurs as volume and not a voxel, hence it is necessary to define the effective volume. The effective volume is the volume around the activated voxel. By selecting an appropriate radius (2 to 5 voxels), the spherical region around the voxel is defined as the effective volume. A fading sphere of desired size is multiplied with time course to generate the volumetric data to be embedded into the baseline spatial data. Appropriate locations are chosen based on the previous ICA analysis data. The following locations (MNI Coordinates) are chosen for embedding the components into the spatial data.

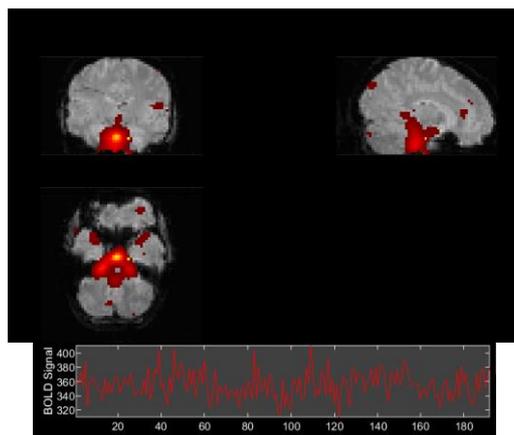
	X	Y	Z
C1	-39.5	26.5	-6.5
C2	0.5	67.5	10.5

C3	0.5	-66.5	-12.5
C4	3.5	-63.5	-27.5
C5	-47.5	43.5	-12.5
C6	3.5	52.5	49.5

Table 2. Spatial location of the components

RESULTS

The result of implementation explained is shown in figure 5 to 8. The methodology to generate simulated fMRI dataset for normal and ADHD subject was implemented using Matlab framework. All the implementation explained above was done on a PC running Microsoft Windows operating system. Pre-processing steps, realignment, normalisation and smoothing was done using SPM12 toolbox running Matlab 2018a. Baseline volume generation (spatial component), time course (temporal component) generation and their fusion were implemented in Matlab. The validation of the simulated dataset was done using the GIFT toolbox. The most commonly used method for analysis of fMRI dataset is the Independent Component Analysis (ICA). ICA is shown to reliably extract the DMN and functional connectivity of the brain from fMRI data[27][28]. Hence we use ICA for testing the simulated dataset. Infomax algorithm[29] available with the GIFT toolbox was applied to real and simulated fMRI data for both normal and ADHD subject. The ICA was able to extract all the components from the simulated fMRI dataset. The activated region is highlighted with a heat map. The consistency of the results was verified by repeating the procedures with different sets of components. Each time the components were extracted by the Infomax algorithm. A complete 4D volumetric resting-state fMRI dataset is simulated within the framework of Matlab with known ground truth. A R-package, neuRosim[2] is a collection of data generation functions for fMRI. For a comparison with our method, it has options to generate time series to 4D volume fMRI dataset. However, it is limited to generation of simulated fMRI volumes for evoked/task-based studies. A time series data generation function simTSrestingstate can be used to generate the resting state time courses. But the package does not have option to integrate the generated resting state time course into the 4D dataset. In our work, we explain a simple and easily implementable method to simulate the resting state fMRI dataset using Matlab. We generate the resting state time series and integrate it into the 4D volume fMRI dataset. The dataset is validated by comparing the ICA results of simulated dataset with real data. From the New York dataset, a normal and ADHD subject data is chosen. Using the baseline volume data, time-courses are embedded to generate the complete volumetric data. With four single subject data available, single component from each dataset was chosen based on similar dynamic range and its fALFF and number of activated voxels were compared. The results of comparison are shown in table 3. Dynamic range is the difference between the peak power and minimum power at frequencies to the right of the peak. Since the simulated time courses were band limited, dynamic range of the components was used as the criteria to select the components for comparison. fALFF is the Low frequency to high frequency power ratio. Activated voxel count is the number of active voxel for the corresponding component.



Component from real data

Figure 5. Orthographic view of the simulated data and its corresponding component from normal subject

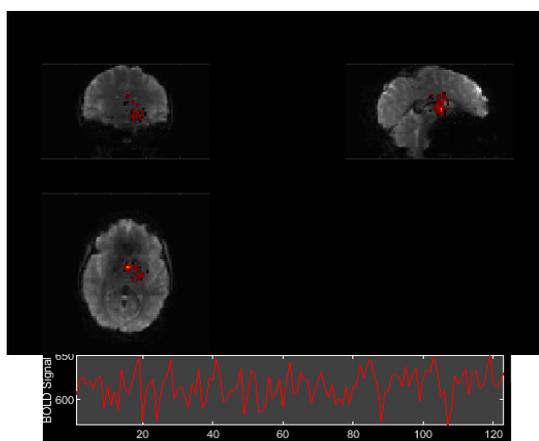


Figure 6. Orthographic view of the simulated data and its corresponding component for normal subject

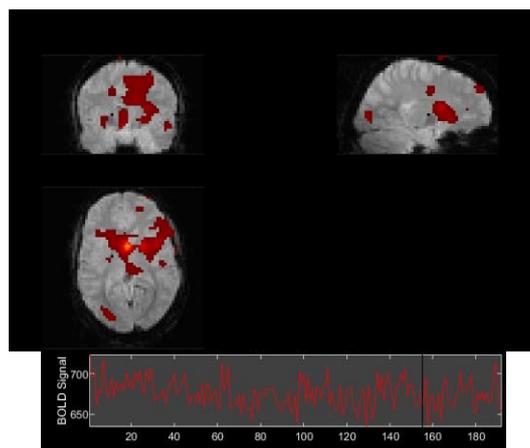


Figure 7. Orthographic view of the real data and its corresponding component for ADHD subject

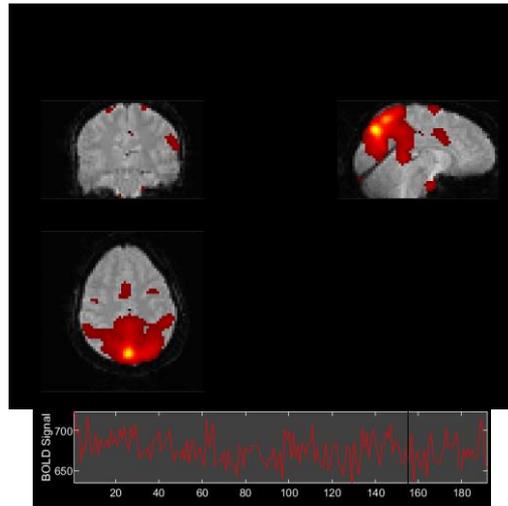


Figure 8. Orthographic view of the simulated data and its corresponding component

Figure 5 and 6 shows orthographic view of fMRI and its corresponding component of real and simulated data for normal subject respectively. Similarly Figure 7 and 8 shows orthographic view of fMRI and its corresponding component of real and simulated data for ADHD subject respectively. From the visual comparison of the data, it can be observed that the simulated fMRI data is similar to the real data. For a parametric comparison of the simulated data with the real data, components with similar dynamic range were chosen. fALFF and the activated voxel count were compared for such components. The summary of the results are shown in table 3.

	Dynamic Range	fALFF	Activated Voxel count
Real Normal	0.059	12.26	84759
Simulated Normal	0.055	8.32	60898
Real ADHD	0.046	24.26	105463
Simulated ADHD	0.047	25.72	116845

Table 3. Comparison of results

From the table 3, it can be observed that for components with similar dynamic range, the fALFF and activated voxel count for the simulated data in normal subject is slightly lesser compared to the real data. For the ADHD subject, the fALFF value and the activated voxel count is similar to the real data. The advantage of having a simulated dataset is that the components are known. Any analysis method that is applied on the real data can be first tested with the simulated dataset to validate the method. The interpretation of the result of such method is reliable as the analysis was done on a known dataset.

CONCLUSION AND FUTURE WORK

A simulated resting-state fMRI dataset has been generated for normal and ADHD subject using Matlab. Most fMRI datasets available from the researchers are event/task-based. Although real resting-state fMRI datasets are available, the ground truth in the data is not

known. Hence there is some uncertainty about the results. The uncertainty can be done away with by validating the analysis method with known data. In this method, we have simulated a complete 4D resting-state fMRI dataset. A real resting-state fMRI dataset from the NITRC website is used for baseline volume data (SM). The voxels preprocessed and averaged over time points. Time courses (TC) are generated using inverse Fourier transform in the frequency band of 0.01 to 0.1 Hz. Noise sources are modeled and added to the time course to mimic the real data. The generated time course is blended with the spatial maps at appropriate locations to complete the simulated resting state dataset. This dataset can be used to validate the analysis methods used for RS-fMRI data.

MATLAB with toolbox, such as SPM12 and GIFT were used for preprocessing, implementation and testing. The simulated resting state fMRI dataset is tested using the Infomax algorithm available in GIFT toolbox. All the six components were extracted by the algorithm. By repeating the procedure many times, duplicates of the actual components were detected a few times in the adjoining areas. The method still has a scope for improvement. The problem of duplicate components must be addressed. Also the noise models of the fMRI data are very complex. Newer models are being proposed which can be included for the future generation of fMRI dataset. Another future work can be in the generation of a complex-valued fMRI dataset. This method can be standardized for the future generation of resting-state fMRI datasets for different pathological conditions by defining a specific framework for each condition.

REFERENCES

- [1] M. Welvaert and Y. Rosseel, "A Review of fMRI Simulation Studies," *PLoS One*, vol. 9, no. 7, p. e101953, Jul. 2014, doi: 10.1371/JOURNAL.PONE.0101953.
- [2] M. Welvaert, J. Durnez, B. Moerkerke, G. Verdoolaege, and Y. Rosseel, "neuRosim: An R package for generating fMRI data," *J. Stat. Softw.*, vol. 44, no. 10, pp. 1–18, 2011, doi: 10.18637/jss.v044.i10.
- [3] J. E. Hill, X. Liu, B. Nutter, and S. Mitra, "A task-related and resting state realistic fMRI simulator for fMRI data validation," in *Medical Imaging 2017: Image Processing*, Feb. 2017, vol. 10133, p. 101332N, doi: 10.1117/12.2254777.
- [4] M. Fukunaga *et al.*, "Large-amplitude, spatially correlated fluctuations in BOLD fMRI signals during extended rest and early sleep stages," *Magn. Reson. Imaging*, vol. 24, no. 8, pp. 979–992, Oct. 2006, doi: 10.1016/j.mri.2006.04.018.
- [5] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *Proc. Natl. Acad. Sci.*, vol. 98, no. 2, pp. 676–682, Jan. 2001, doi: 10.1073/PNAS.98.2.676.
- [6] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nat. Rev. Neurosci.*, vol. 8, no. 9, pp. 700–711, Sep. 2007, doi: 10.1038/nrn2201.
- [7] L. EC *et al.*, "Integration of resting state functional MRI into clinical practice - A large single institution experience," *PLoS One*, vol. 13, no. 6, Jun. 2018, doi: 10.1371/JOURNAL.PONE.0198349.

- [8] T. C, W. Y, Z. J, and N. J, “Different Developmental Pattern of Brain Activities in ADHD: A Study of Resting-State fMRI,” *Dev. Neurosci.*, vol. 40, no. 3, pp. 246–257, Oct. 2018, doi: 10.1159/000490289.
- [9] N. A. Khan, S. A. Waheeb, A. Riaz, and X. Shang, “A Novel Knowledge Distillation-Based Feature Selection for the Classification of ADHD,” *Biomolecules*, vol. 11, no. 8, p. 1093, Jul. 2021, doi: 10.3390/BIOM11081093.
- [10] M. D. Figueroa-Jimenez *et al.*, “Complexity analysis of the default mode network using resting-state fmri in down syndrome: Relationships highlighted by a neuropsychological assessment,” *Brain Sci.*, vol. 11, no. 3, pp. 1–19, Mar. 2021, doi: 10.3390/BRAINSKI11030311.
- [11] C. Shi, J. Zhang, and X. Wu, “An fmri feature selection method based on a minimum spanning tree for identifying patients with autism,” *Symmetry (Basel)*, vol. 12, no. 12, pp. 1–15, Dec. 2020, doi: 10.3390/SYM12121995.
- [12] W. K. W. Lau, M.-K. Leung, and B. W. M. Lau, “Resting-state abnormalities in Autism Spectrum Disorders: A meta-analysis,” *Sci. Reports 2019 91*, vol. 9, no. 1, pp. 1–8, Mar. 2019, doi: 10.1038/s41598-019-40427-7.
- [13] N. A. Khan, S. A. Waheeb, A. Riaz, and X. Shang, “A three-stage teacher, student neural networks and sequential feed forward selection-based feature selection approach for the classification of autism spectrum disorder,” *Brain Sci.*, vol. 10, no. 10, pp. 1–22, Oct. 2020, doi: 10.3390/BRAINSKI10100754.
- [14] V. P, J. DT, and J. CR, “Resting state functional MRI in Alzheimer’s Disease,” *Alzheimers. Res. Ther.*, vol. 4, no. 1, 2012, doi: 10.1186/ALZRT100.
- [15] J. Lu *et al.*, “Functional connectivity between the resting-state olfactory network and the hippocampus in Alzheimer’s disease,” *Brain Sci.*, vol. 9, no. 12, Dec. 2019, doi: 10.3390/BRAINSKI9120338.
- [16] N. V. Kraguljac, A. Srivastava, and A. C. Lahti, “Memory Deficits in Schizophrenia: A Selective Review of Functional Magnetic Resonance Imaging (fMRI) Studies,” *Behav. Sci. 2013, Vol. 3, Pages 330-347*, vol. 3, no. 3, pp. 330–347, Jun. 2013, doi: 10.3390/BS3030330.
- [17] S. Li *et al.*, “Dysconnectivity of Multiple Brain Networks in Schizophrenia: A Meta-Analysis of Resting-State Functional Connectivity,” *Front. Psychiatry*, vol. 0, no. JULY, p. 482, 2019, doi: 10.3389/FPSYT.2019.00482.
- [18] W. AF, van de W. SCF, L. AFG, D. AA, J. HIL, and K. ML, “Resting-state fMRI in Parkinson’s disease patients with cognitive impairment: A meta-analysis,” *Parkinsonism Relat. Disord.*, vol. 62, pp. 16–27, May 2019, doi: 10.1016/J.PARKRELDIS.2018.12.016.
- [19] P. Tuite, “Brain magnetic resonance imaging (MRI) as a potential biomarker for parkinson’s disease (PD),” *Brain Sci.*, vol. 7, no. 6, Jun. 2017, doi: 10.3390/BRAINSKI7060068.

- [20] S. . Pekar, J.J. Mostofsky, “FCP Classic Data Table.” http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html.
- [21] N. H. Yuen, N. Osachoff, and J. J. Chen, “Intrinsic Frequencies of the Resting-State fMRI Signal: The Frequency Dependence of Functional Connectivity and the Effect of Mode Mixing,” *Front. Neurosci.*, vol. 0, p. 900, Sep. 2019, doi: 10.3389/FNINS.2019.00900.
- [22] X. Bu *et al.*, “Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder,” *Transl. Psychiatry 2019 91*, vol. 9, no. 1, pp. 1–10, Jan. 2019, doi: 10.1038/s41398-018-0362-9.
- [23] C. G. Fox, “An inverse fourier transform algorithm for generating random signals of a specified spectral form,” *Comput. Geosci.*, vol. 13, no. 4, pp. 369–374, Oct. 1987, doi: 10.1016/0098-3004(87)90009-4.
- [24] T. T. Liu, “Reprint of ‘Noise contributions to the fMRI signal: An Overview’,”” *Neuroimage*, vol. 154, pp. 4–14, Jul. 2017, doi: 10.1016/J.NEUROIMAGE.2017.05.031.
- [25] “SPM12 Software - Statistical Parametric Mapping.” <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/> (accessed Jul. 25, 2021).
- [26] V. D. Calhoun, “Group ICA Of fMRI Toolbox(GIFT).” <https://trendscenter.org/software/gift/> (accessed Jul. 25, 2021).
- [27] V. Kiviniemi, J. H. Kantola, J. Jauhiainen, A. Hyvärinen, and O. Tervonen, “Independent component analysis of nondeterministic fMRI signal sources,” *Neuroimage*, vol. 19, no. 2, pp. 253–260, Jun. 2003, doi: 10.1016/S1053-8119(03)00097-1.
- [28] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, “Investigations into resting-state connectivity using independent component analysis,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 360, no. 1457, pp. 1001–1013, 2005, doi: 10.1098/RSTB.2005.1634.
- [29] X. Wu, L. Yao, Z. Long, and H. Wu, “Improved Infomax algorithm of independent component analysis applied to fMRI data,” *Med. Imaging 2004 Image Process.*, vol. 5370, pp. 1880–1889, May 2004, doi: 10.1117/12.534673.